Journal of Applied Psychology

JOHN G. DARLEY, Editor
University of Minnesota

Table of Contents

Guttman-Type Scales for Union and Management Attitudes Toward Each Other: R. Stagner, W. E. Chalmers, and M. Derber 293

The Attitudes of West Texas College Students Toward School Integration: H. Greenberg and Dolores Hutto 301

Anxiety Level and Score on a Biographical Inventory: D. H. Kausler and E. P. Trapp 305

Development of a Criterion of Research Productivity: R. E. Stoltz 308

Performance as a Function of Control-Display Relations, Positions of the Operator, and Locations of the Control: M. Humphries 311

The Postwar Occupational Adjustment of Emotionally Disturbed Soldiers: J. B. Miner and J. K. Anderson 317

Adequacy of the Residual Sensory Cues for Psychomotor Performance of Arm Amputees: Hilde Groth and J. Lyman 323

Curriculum Assessment with Critical Incidents: A. S. Glickman and T. R. Vallance 329

A Personality Inventory Employing Occupational Titles: J. L. Holland 336

Length of Work Periods in Visual Research: M. A. Tinker 343

Communication Restraints, Group Flexibility, and Group Confidence: R. C. Ziller 346

The Comparative Effectiveness of Some Psychological and Physiological Measures in Ranking the Impact of Diverse Environmental Conditions: B. J. Fine 353

American Psychological Association

Volume 42, Number 5 October, 1958

Consulting Editors

HAROLD E. BURTT, *Ohio State University*
ALPHONSE CHAPANIS, *Johns Hopkins University*
CLIFFORD E. JURGENSEN, *Minneapolis Gas Company*
LAURENCE S. MCGAUGHRAN, *University of Houston*
QUINN MCNEMAR, *Stanford University*
ALEXANDER MINTZ, *City College of New York*
HAROLD F. ROTHE, *Fairbanks, Morse and Company*
JULIAN B. ROTTER, *Ohio State University*
THOMAS A. RYAN, *Cornell University*
DONALD E. SUPER, *Columbia University*
MILES A. TINKER, *University of Minnesota*
ALFRED C. WELCH, *University of New Mexico*

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
SARAH WOMACK, Editorial Assistant

This journal gives primary consideration to original investigations in any field of applied psychology except clinical and consulting psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: vocational and educational prognosis, diagnosis, and guidance at the secondary and college level; personnel research in business, industry, and government; biomechanics; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research.

Because of the large number of manuscripts submitted, authors should adhere to the rule of "brevity consistent with clarity." The typical manuscript should run to approximately 4,000 words. There is a lag of approximately twelve months between receipt and publication of an article. Authors may request advanced publication if they are prepared to pay the cost of printing the necessary extra pages.

Manuscripts should be addressed to the Editor, John G. Darley, 408 Johnston Hall, University of Minnesota, Minneapolis 14, Minnesota. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies.

Manuscripts must conform to the style requirements described in the *Publication Manual of the American Psychological Association.*

Journal of Applied Psychology

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth Street N.W.
Washington 6, D. C.

$8.00 per volume $1.50 per issue

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized October 10, 1947.

© 1958 by the American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 42, NO. 5 OCTOBER, 1958

Guttman-Type Scales for Union and Management Attitudes Toward Each Other [1]

Ross Stagner

Wayne State University

W. E. Chalmers and Milton Derber

University of Illinois

The importance of "the attitudes of the parties toward each other" has long been recognized in industrial relations. Many writers have used the concept, with and without reference to specific indices of attitude. However, so far as we know, no one has done any systematic work on the preparation of scales for the quantitative measurement of this variable. Such scales as have been published (e.g., in the Illini City studies, [4]) have been arbitrarily put together and have not been subjected to scaling analysis.[2] Many of these scales have included specific items referring to local situations, hence were not useful in comparative studies across establishments.

For the above reasons, it seemed desirable to construct generalized scales which could be presented to management and union officers for the purpose of obtaining a quantified statement of attitude. The substantive context within which the scales were to be used is reported in other publications (see, e.g., 3). The present report is therefore limited to the technical characteristics of the scales themselves.

Selection of scale items. Since it was our desire to prepare a generalized scale covering the major elements which might be involved in the "attitude of management toward the union" or in the "attitude of union toward management," we selected items from previous scales, and from discussions in the literature, which seemed likely to have such generalized applicability. Thus, in the scale for unions, items were included about management policies, about formen, middle and top management, about managerial use of power, and so on. Parallel questions were framed to be presented to executives for attitude toward the union.

Respondents. The scales presented here are apparently unique also in another respect. They are based upon pooled answers from two respondents, instead of treating each respondent as a separate member of the population. This was an important component of our research design. Data regarding each establishment (with few exceptions) were based on the pooled answers of the two top union officials, those who bargained directly with the company, and on the pooled answers of the two top executives who handled union relations for management. This procedure was followed to increase reliability and to decrease the role of subjective biases in the perception of situations.[3] Specifically, in the case of the

[1] We should like to express our appreciation to Milton Edelman of Southern Illinois University and to former graduate assistants, Herbert Schaeffer, Robert Ver Nooy, Robert Mitchell, Sheldon Luskin, and John Tipton, who aided in various phases of the collection and analysis of the data.

[2] It should perhaps be noted that the items used in the Illini City studies with rank-and-file employees were analyzed by Williams (5) who found that both the attitude-to-company and attitude-to-union items formed acceptable scales. However, these were not used with union or company officials.

[3] For example, there were numerous disagreements between our pairs of managerial respondents even on "factual" issues such as seniority practice. The use of two informants presumably gave us more "objec-

attitudinal material, we wanted a figure which represented "the position of management," not a single person's attitude, and similarly for the union.

The analysis reported here is based upon an *N* of 41, meaning that interviews were conducted in 41 establishments in three downstate Illinois communities. Actually, 76 executives were interviewed (in 6 companies, only one executive could be found who was actively involved in union relations). Similarly, the total number of interviews was 81 for union officials; in all but one case, two men were interviewed to obtain the union position. Interviews were conducted in plant or union offices by our field interviewers. Each item with its response alternatives was presented, and the interviewer marked the response selected. Respondents had no opportunity to consult with each other regarding answers to these items.

The population of 41 establishments has been described elsewhere (1). Suffice it to say that the range of sizes (hourly paid employees) was from 73 to 2,100 employees; some were in manufacturing, others in utilities and services; among the manufacturers, both producer and consumer goods were involved; and a wide variety of unions functioned as bargaining agents. However, these 41 firms were restricted to cases which were not branches of larger organizations, and to cases in which a single union bargained for a majority of the work force.

Since each item was presented as a four- or five-alternative multiple-choice question, pooling was accomplished simply by taking the average of the positions chosen. Perfect agreement between the two spokesmen was not common. For managers, agreement was found on five items in at least 50% of firms. For unionists, seven items showed agreement in at least 50% of the cases. Since 14 items were used, it is clear that neither "the management position" nor "the union position" is clearly established and unequivocally stated in the typical relationship. However, it should be noted that differences in response were usually limited to a single step on the answer scale.

Averaging the responses of two persons implies the assumption that the answer steps represent equal intervals. This assumption has not been tested. Since the final product is presumed to be only an ordinal scale, some inequality of answer steps would seem not to introduce a serious question.

Procedures

The task of this analysis, in generalized terms, is to determine the functional unity of the specific attitude items, i.e., to test the hypothesis that there is a single point on a hypothetical continuum which correctly represents the position of one party's attitude with respect to the other. This would be in contrast to an equally plausible finding, viz., that each attitude is truly composed of a number of independent components, and that the selection of an index number to designate one party's position is merely a case of averaging positions on specific issues not intrinsically related one to another. Research on union-management relations has commonly assumed that these attitudes are unitary; we felt it important to test this assumption.

Two methods are commonly used for establishing the unitary nature of a function revealed in a series of discrete responses: correlational analysis; and scaling technique. For purposes of cross-comparisons, we used both. Phi coefficients were computed between all pairs of items, to permit determination of clusters of related items; and Guttman's scalogram analysis (3) was also applied to the data. The following items were used in the analysis (the items are ranked as in the final scale; marginal numbers indicate position in original questionnaire; numbers in parentheses, such as (2.5), indicate the cutting point between an answer considered favorable and one considered unfavorable):

Scale for Management Attitude Toward Union

43. Are the union officers effective leaders of their organization?
    (1)——very much so
    (2)——pretty good
    (3)——mediocre
    (4)——very poor (2.5)

tive" data than reliance on a single person, especially since we reinterviewed on "factual" questions in case of disagreement. No reinterviews were conducted on attitudinal items.

35. Is the union generally reasonable or not in its claims?
    (1)——very reasonable
    (2)——reasonable most of the time
    (3)——frequently unreasonable
    (4)——extremely unreasonable (2.5)

36. Does the union interfere seriously with how the company is managed, or does the management have a reasonably free hand in running the plant?
    (1)——union is no problem
    (2)——it interferes a little but not seriously
    (3)——it interferes quite often
    (4)——it seriously interferes with management (2.0)

40. Are the union officers interested in the welfare of the rank-and-file workers?
    (1)——very much so
    (2)——pretty much
    (3)——slightly
    (4)——very little (2.0)

38. Does the union cooperate with management on production matters or not?
    (1)——they are extremely cooperative
    (2)——they will go along but not positively support
    (3)——they do not interfere seriously but sometimes are obstructionist
    (4)——they restrict production improvements quite often (2.5)

34. In general, how do you personally feel about your company's relations with the union?
    (1)——very satisfied
    (2)——moderately satisfied
    (3)——moderately dissatisfied
    (4)——very dissatisfied (2.0)

39. Has the union tended to weaken employee discipline, or has it cooperated with management on disciplinary matters?
    (1)——cooperative and helpful
    (2)——sometimes helps but not always
    (3)——sometimes interferes with discipline
    (4)——has created some serious disciplinary problems (2.5)

46. Does the union have too much power in your establishment?
    (1)——not too much
    (2)——too much in a few respects
    (3)——too much in many respects
    (4)——far too much (1.5)

44. Does the union have the support of the workers?
    (1)——most of the workers are strongly behind it
    (2)——only a few really active people but most workers go along
    (3)——not too much feeling either way
    (4)——a lot of the workers are hostile (1.0)

37. How do you feel about using the union as the main channel of communication to the workers on company policies?
    (1)——strongly favor
    (2)——moderately favor
    (3)——moderately oppose
    (4)——strongly oppose (combined with 41)

41. Are the *local* union officers skillful bargainers?
    (1)——very much so
    (2)——pretty good
    (3)——mediocre
    (4)——very poor (favorable if $37 + 41 = 3$ or less)

The following items did not scale:

42. Are the international union representatives skillful bargainers?
    (1)——very much so
    (2)——pretty good
    (3)——mediocre
    (4)——very poor
    (5)——none are involved

45. Do the international union officers create any serious problems or not?
    (1)——they are generally responsible and helpful
    (2)——they are more helpful than troublesome
    (3)——they are more troublesome than helpful
    (4)——they generally stir up trouble
    (5)——none are involved

47. Does the union try to live up to its agreements?
    (1)——always
    (2)——usually
    (3)——frequently does not
    (4)——rarely

Scale for Union Attitude Toward Management

43. Are the top management officials effective executives of the establishment?
    (1)——very much so
    (2)——pretty good
    (3)——mediocre
    (4)——very poor (2.0)

44. What is the top management attitude toward the union?
    (1)——strongly favorable
    (2)——moderately favorable
    (3)——moderately unfavorable
    (4)——strongly unfavorable (2.0)

47. Does the company try to live up to its agreements?
    (1)——always
    (2)——usually
    (3)——frequently does not
    (4)——rarely (2.0)

46. Does the company abuse its power in this establishment?
    (1)——rarely
    (2)——occasionally
    (3)——frequently
    (4)——very often (1.5)

34. In general, how do you personally feel about your union's relations with the company?
    (1)——very satisfied
    (2)——moderately satisfied
    (3)——moderately dissatisfied
    (4)——very dissatisfied (2.0)

36. Has the management shown any understanding of your problems as a union officer?
    (1)——very understanding
    (2)——understands the union situation pretty well
    (3)——understanding of union problems is limited
    (4)——little or no understanding of union problems (2.0)

37. Has the management tried to undermine the union position through direct dealings with the workers, or has it been careful to safeguard the union position in such contacts?
    (1)——is always careful not to hurt union
    (2)——is usually careful not to hurt union
    (3)——occasionally tries to weaken union
    (4)——frequently tries to weaken union (2.0)

35. Is the top management generally reasonable or not when it comes to discussing union claims?
    (1)——very reasonable
    (2)——reasonable most of the time
    (3)——frequently unreasonable
    (4)——extremely unreasonable (1.5)

40. Are the top management officials interested in the welfare of the workers?
    (1)——very much so
    (2)——pretty much
    (3)——slightly
    (4)——very little (1.5)

The following items did not scale:

38. Do the foremen, in general, act toward the union in the same way as top management?
    (1)——the foremen are very much easier to get along with
    (2)——the foremen are somewhat easier to get along with
    (3)——the foremen are about the same as top management
    (4)——the foremen are somewhat more difficult to get along with
    (5)——the foremen are much more difficult to to get along with

39. Does middle management, in general, act toward the union in the same way as top management?
    (1)——the middle management are very much easier to get along with
    (2)——they are somewhat easier to get along with
    (3)——they are about the same as top management
    (4)——they are somewhat more difficult to get along with
    (5)——they are much more difficult to get along with

41. Are the top management officials skillful bargainers?
    (1)——very much so
    (2)——pretty good
    (3)——mediocre
    (4)——very poor

42. (If a multi-plant company with a home office outside this establishment) Are the company representatives from the home office skillful bargainers?
    (1)——very much so
    (2)——pretty good
    (3)——mediocre
    (4)——very poor
    (5)——not involved

45. (If a multi-plant company with a home office outside this establishment) Do company representatives from the home office create serious problems or not?
    (1)——they are generally responsible and helpful
    (2)——they are more helpful than troublesome
    (3)——they are more troublesome than helpful
    (4)——they generally stir up trouble
    (5)——not involved

Results

1. *Attitude of management toward the union.* The 14 × 14 matrix gives 91 phi coefficients, of which 26 are positive and significantly above zero,[4] while only 13 are negative, one of these being significant.

This finding would suggest the existence of a single common factor. However, inspection of the data indicates that only 7 items are closely related; 15 of their 21 intercorrelations are positive and significant. Of the remaining 70 coefficients, only 11 are positive and significant. Hence, the common variable appears to be defined by such items as 34 and 46 (how do you personally feel about company's relations with union, and does union have too much power). Next are items 36, 39, 40 and 47, followed closely by item 38.

[4] Since the data were dichotomized to compute phi coefficients, significance was tested by finding the 5% value of chi square for $df = 1$. This value, 3.84, corresponds to phi of .293. Copies of tables of phi coefficients may be obtained by writing to M. L. Derber, University of Illinois, Urbana, Illinois.

The remaining seven items have few significant correlations either with the previously listed seven or among themselves. A decision based solely on this set of data would undoubtedly have been to take the seven items with a majority of the significant positive coefficients and use them as a scale, rejecting the remaining seven which show random interrelationships.

The second analysis was based on the Guttman procedure. An arbitrary score was computed for each establishment by summing the values of the answers to the 14 items; and the establishments were ranked in order of favorableness toward the union. Responses to the 14 items were then tallied in 14 columns, with the item showing the largest number of favorable opinions at the left, and decreasing toward the right. Cutting scores were then sought such that a maximum number of favorable responses to a given item would be in the upper portion of the column, and unfavorable responses concentrated in the lower portion.[5] By dropping items, recomputing scores, reranking establishments, etc., it was possible to minimize the number of "errors"—unfavorable responses above this arbitrary line, favorable responses below it.

In the course of this scaling, Items 42 and 45 were dropped first. Item 42 had only one significant positive phi coefficient, and 45 had the single significant negative phi, with no significant positives. The results of the two methods thus far are closely comparable. In a later phase of scaling Item 47 was also dropped. On phi analysis 47 had a large number of significant coefficients (six). We cannot explain why it proved impossible to find a suitable cutting point on 47 which would relate it suitably to the revised total score. Finally, Items 37 and 41 were kept in the scale by combining them; i.e., the choice of answers on the two items had to add to 3 or less to be counted as favorable. Otherwise, the combined response was treated as unfavorable. Thus, the Guttman scaling procedure gives us a scale using 10 items, six being in the list of seven having high intercorrelations. Two items, 35 and 43, were found to scale adequately despite a modest number of significant correlations, and two others, 37 and 41, were saved by combining them to make one scoring unit, a procedure which is feasible with Guttman scaling but difficult to achieve by straight correlational methods.[6]

Guttman determines whether he has a "true scale" by computing the "coefficient of reproducibility." This is simply the number of responses correctly located as favorable or unfavorable (see above) divided by the total number of responses. For the final management scale, this coefficient is .915, indicating the presence of a "true scale" in this case.

Each item in a Guttman scale is scored as having a weight of one point. Thus, there is a possible range of scores from zero to nine. These are the scores which have been used in the substantive analysis of our data.

2. *Union scale.* We may now consider the parallel scaling of the responses obtained from union respondents. It will be noted that the items presented to the union spokesmen were in general similar to those for management, but certain variations were necessary.

The phi coefficient analysis for the 14 union items indicates that the degree of internal coherence of attitudes here is somewhat less than for the company officials; only 20 coefficients (out of 91) are positive and significant at the 5% level; again, one is negative and significant to this degree.

The common thread in the union items seems best represented by Item 44 (what is top management attitude to union) and Item 46 (does company abuse power), which have significant correlations with six other items. Close behind are 35, 36, and 40 (is top management reasonable, has management shown understanding, top management interested in worker welfare), each of these having five significant correlations. Reasonably integrated with the others is Item 34 (how do you personally feel about union relations with company). One item, 38, had no significant coefficients, and Items 41 and 42 had only one each. It would appear that a satisfactory

[5] For precise details of the Guttman method, see (3).

[6] Guttman defends the procedure of contriving these combined items on the ground that a pattern of responses may provide information not given by an isolated response.

scale could be prepared with Items 44, 46, 35, 36, 40, and 34, since 14 of the 20 significant positive coefficients are found within this group; only one of their intercorrelations was below the 5% cutting point.

Turning now to the Guttman analysis, we find a situation similar to that in the management data. Four items were dropped at the first stage of the analysis: 38, 39, 42, and 45; collectively, these four items accounted for only three of the 20 significant positive coefficients, and for the single negative coefficient which reached significance. It would appear, therefore, that items which correlate very poorly with others in the pool are fairly certain to drop out early in a Guttman analysis.

After rescoring and reranking the establishments, we found it necessary to drop one more item, 41; this item likewise has only a single significant correlation. The remaining nine items were shown to form a "true scale"; the coefficient of reproducibility was .902. It is interesting to note that the item with the fewest "errors" in the final scaling is 44, which was also best in the correlational data; but, on the other hand, Item 46, almost as good in the correlation matrix, has almost the highest number of errors in the Guttman procedure. Since similar anomalies were observable in the management data, we infer that a core of items established by intercorrelations will agree fairly well with those selected in the early stages of the Guttman method; but as further precision is sought, the correlational results do not predict well the final scaling of items.

Guttman considers his scales to be truly "ordinal" in character, and instead of speaking of scores he speaks of "scale types." However, he notes that it does little harm to the data to treat everyone making the same numerical score as belonging to the same scale type, even though the pattern of favorable and unfavorable answers may not be exactly identical for all. We have followed this suggestion for simplicity, and thus have treated all establishments having m favorable answers as belonging to scale type m. Since the number of errors is small, the number of cases which would be reclassified by a different procedure is very small.

Table 1

Distribution of "Scale Types"[a]

	Management to Union	Union to Management
10	3	—[b]
9	0	5
8	13	11
7	11	7
6	1	5
5	3	5
4	3	3
3	0	1
2	4	0
1	2	3
0	1	1

[a] We have treated the scale type simply as the sum of favorable responses, and statistically have considered it as the equivalent of a test score.
[b] The maximum for the union scale is 9.

The final distribution of scale types is shown in Table 1 for both scales. It will be noted that the entire attitudinal range is represented in our population of establishments on each scale; however, there is not a uniform spread along the scale as Guttman would consider desirable. The piling up is in the "more favorable" categories; this is presumably fortunate in terms of labor peace.

An additional technical note is important. Ideally, in Guttman scaling, the marginal entries should be widely and fairly uniformly spread; i.e., there should be a range from an item answered favorably by almost everyone to one which is answered unfavorably by most of the population, and items well spaced between these two. Our data do not meet this test; despite our choices of cutting-points, the items also pile up in the range with large numbers of favorable answers. This situation could not be corrected without seriously increasing the number of "errors." The errors remaining in the final scales were randomly distributed.

3. *"Attitudinal climate."* In our Illini City study, the data indicated that a common "attitudinal climate" could be identified for an establishment. This conclusion was based primarily on the responses of rank-and-file workers (who were also union members), indicating that liking for the company was associated

with liking the union, and vice versa. In a plant where the workers liked management, they seemed also favorable to the union; in firms where the management was disliked, the union also was in relative disfavor.[7] However, this relationship did not hold when we compared the attitudes of union and management leaders within the establishment.

In our present study, it is possible to make another test of the "attitudinal climate" hypothesis. We have used reasonably parallel questions, asked of managers and of unionists, to define the attitude of each toward the other. If attitudinal climate generalizes to all levels of an establishment, we would expect that firms in which management approves of the union would also have a union which approves of management; i.e., there should be a significant positive correlation between the two measured attitudes.

This prediction is not confirmed. The correlation between the two attitudes, for our 41 establishments, using the a priori scales based on the original 14 items in each questionnaire, was .271. This verged on significance, and it was thought that the refined scales would correlate more highly. Just the reverse proved to be the case. The final scales correlated only .130 with each other, a result which would indicate that there is no connection, except perhaps in very extreme cases, between the two attitudes.[8] We are therefore disposed to conclude that, while attitudinal climate may be a functional reality at the level of the rank and file, the relationship at the level of the leaders is too complex to be incorporated in a simple formulation of this kind.

[7] Compare the recent study by Form and Dansereau (2), in which it was shown that active union members like their jobs and their departments ("as a place to work") more than inactive members.

[8] A substantive analysis indicates that there may be a variety of reasons for management to report hostility to the union, and similarly on the part of the union. Certain combinations of these factors seem to give rise to "types" or patterned relationships which are repeated several times in our sample of 41 establishments. It should be noted that, in constructing these "types," we did not assume either a high correlation of management and union attitudes or the legitimacy of a combined attitude score. Types were constructed by choosing a cutting point on each scale separately and treating as a single type those establishments which had certain combinations of high or low scores (see [1] for details).

Discussion

This report deals with a portion of the data from an extensive quantitative study of union–management relations at the local level (1). A Guttman analysis (supported by correlational data) indicates that unidimensional scales have been established both for management attitudes to the union and for union attitudes toward management.

The scales developed are strongly characterized by what may be called a "halo" factor, or a generalized tendency to approve or disapprove. The better items in the management scale were such as, "How do you personally feel about the company's relations with the union?" "Does the union have too much power?" However, this does not mean the respondent approved (or disapproved) of everything associated with the union. Highly specific items failed to scale; e.g., "Are the international union representatives skillful bargainers?" The same tendency (for generalized items to scale, and for specific items not to scale) occurs with the union attitudes. It would appear that a Guttman scale may be simply an elaborate and highly reliable way to ask, "Do you approve of the union?"

Such scales can, of course, be of considerable value in research which involves quantitative estimates of union or management attitudes. It is to be hoped that the scales here reported will be seen by other students of industrial relations as valuable devices, to be incorporated in their own research designs. This seems to be a path by which progress in this field can be greatly accelerated.

Summary

1. Fourteen multiple-choice items, relating to the attitude of management toward the union, were presented to the two top labor-relations officials of 41 Illinois companies. Correlational analysis indicated the presence of a single factor, or unitary attitudinal dimension, involving at least seven items. Guttman analysis confirmed this, and produced a "true scale" consisting of nine of the original items plus one contrived item combining two questions.

2. A parallel analysis of responses from the union officials in these same establishments

led to similar conclusions and to a Guttman scale of nine of the original items.

3. Attitude of management toward union is uncorrelated with attitude of union toward management, when measured by these refined scales.

4. The refined scales are valuable instruments for substantive research on union-management relations at the plant level.

Received January 17, 1958.

References

1. Derber, M., Chalmers, W. E., & Stagner, R. Uniformities and differences in local union-management relationships. *Indus. Labor Rel. Rev.*, 1957, 11 (Oct.), 56–71.
2. Form, W. H., & Dansereau, H. K. Union member orientation and patterns of social integration. *Indus. Labor Rel. Rev.*, 1957, 11 (Oct.), 3–12.
3. Guttman, L. The basis for scalogram analysis. In: S. A. Stouffer et al., *Measurement and prediction*. Princeton, N. J.: Princeton Univer. Press, 1950. Pp. 60–90.
4. Institute of Labor and Industrial Relations, University of Illinois. *Labor-management relations in Illini City*. Vol. 1, *Case studies*. Vol. 2, *Exploration in comparative analysis*. Champaign, Ill.: Author, 1953, 1954.
5. Williams, L. K. Scaling of interview data on two union-management relationships. M.A. thesis, Univer. Illinois, 1954.

Journal of Applied Psychology
Vol. 42, No. 5, 1958

The Attitudes of West Texas College Students Toward School Integration

Herbert Greenberg and Dolores Hutto

Texas Technological College

With the recent Supreme Court decision making segregation in schools unconstitutional, interest is high in the question of school integration. Much has been said about the consequences of integration if effected too rapidly, and on the other side, the pressure has been great to desegregate just as quickly as the physical change could be effected. Few attempts have been made, however, to scientifically determine the attitudes of the students themselves concerning integration. Thus, the senior author and his associates conducted a study (2) of Negro and white high school students in the West Texas area. A logical next step it seemed was to conduct a similar study on the college level. This seemed especially appropriate after the incidents in Alabama and the contrary smooth integration at the University of North Carolina.

Problem

1. To determine the attitudes of students in a large West Texas college toward school integration in general and to specific areas of this integration

2. To determine whether there is a relationship between authoritarianism and attitudes toward school integration

3. To determine whether classification in school is related to authoritarianism or to integration attitudes

Procedures

Population. A battery of tests was administered to 460 students of Texas Technological College of Lubbock, Texas. This number represents approximately 7 per cent of the student population and was selected on the basis of an exact breakdown as to division, i.e., engineering, liberal arts, etc.; classification; and sex. The tests were administered under the supervision of the senior author by administrators to seven elementary psychology classes. The personal data were then checked to determine the divisional, classification, and sex breakdown obtained from the sample. Where the test sample deviated from correct representation of the universe, i.e., student population of Texas Tech, cases were eliminated from the over-represented groups till the tested population did accurately represent the universe. This could successfully be done for sex and divisional breakdown, but would have required an overly severe paring in order to create a proportionately representative group, from the senior and graduate classes, as comparatively few of these students could be found in elementary psychology. To make up this deficiency, students were approached at random in the cafeteria and asked their classification and division, and those students who fitted the need were asked if they would cooperate by filling out a short questionnaire. Agreement was received in every case, so that the problem of selective sampling need not be considered. Through this method, the exact categorical breakdown discussed above could be achieved while allowing the bulk of the sampling to be derived from group testing in the seven psychology classes.

The test battery. The California F Scale, the Integration Attitude Scale and a personal data form were employed. The California F Scale (1) was designed to measure authoritarianism in personality through a series of 29 statements to which the respondent reacts by strongly agreeing, agreeing, mildly agreeing, strongly disagreeing, and mildly disagreeing. An example item is "Familiarity breeds contempt."

The Integration Attitude Scale (2), "IA," was designed to measure attitudes toward school integration in general and specific areas in integration in particular. Thus, it is the purpose of this instrument not merely to obtain over-all integration attitudes, but to determine integration attitudes as applied to sports, classrooms, eating facilities, etc. For the purposes of this present study, question Number 24, "I do not think I would be willing to sit next to a member of another race in class," was omitted and was replaced by, "I would not accept a member of the other race as a roommate in the dormitory."

The personal data form was designed to secure personal-social data so as to permit the establishment of comparison groups between the classification, sexes, and divisions leaving integration attitudes and authoritarianism as variables.

Procedures in treating data. Means, standard deviations were found for each classification, division, and sex for each scale. Critical ratios were then obtained between the means of each of these breakdowns for each scale. Thus, freshmen F-scale means were compared with those of the sophomores, juniors, seniors, and graduates while the same procedure was followed with the IA. Likewise, comparisons were

Table 1

Product-Moment Correlation, IA to F

	Product-Moment Correlation	*N*
Freshmen	0.215	134
Sophomores	0.167	124
Juniors	0.15	107
Seniors	0.289	68
Graduates	0.518	23
Total Population	0.213	456

Table 2

Means and Standard Deviations for the Several Classifications for Both Scales

Classification	Mean F	*SD* F	Mean IA	*SD* IA	*N*
Freshmen	108.83	20.008	91.78	41.09	134
Sophomores	108.59	21.66	94.63	11.27	124
Juniors	113.26	21.64	93.63	37.82	107
Seniors	103.7	22.3	98.13	40.58	68
Graduates	105	17.44	97.39	42.5	23

made between males and females for both scales and between the several divisions.

Lastly, an item analysis of the IA Scale was conducted in which the number of positive and negative responses for each item was determined for each classification, division, sex, as well as for the total population. This was done to ascertain group attitudes in specific areas of integration.

Results

Table 1 shows product-moment correlations for the classifications and for the total population IA to F. It can be seen that there is a very low correlation between the two scales, the only significant one being the +0.52 for the graduates. It should be noted further that the interscale correlation computed for the several divisions and for each sex yielded approximately the same results as for the total population. Thus, they are not presented here in separate tables.

Tables 2 and 3 show means and standard deviations for each classification and the critical ratios between designated comparison classifications. It should be noted that a high score designates greater authoritarianism and more negative attitudes toward integration. Thus, the freshmen mean of 108.83 on the F scale makes that group significantly more authoritarian than the seniors, having a mean of 103.7. Conversely, on the IA scale the seniors, with a mean of 98.13, have significantly more negative attitudes toward school integration than the freshmen, whose mean is the lower score of 91.78. It should be noted further that, as in the case of Table 1, there

Table 3

Critical Ratios Between the Means of the Several Classifications for Both Scales

Classification	C.R. "F"	C.R. "IA"
Freshmen/Sophomores	.29	1.92 .05% level
Sophomores/Juniors	1.66	1.87 .05% level
Juniors/Seniors	2.97 .01% level	3.41 .01% level
Freshmen/Seniors	1.73 .05% level	5.26 .001% level
Freshmen/Juniors	1.72 .05% level	1.62
Sophomores/Seniors	1.88 .05% level	2.83 .01% level

was no significant difference between the means of each sex and between the several divisional means. Thus, they are not presented separately.

In most areas the student attitude toward integration is quite positive. For the most part, however, the more advanced classes appear less positive than the freshmen and sophomores. It can be further noted that in the areas of closer and closer personal-social contact the relative positiveness of the answer decreases until such areas as dating (Item 20, "I believe that dating between races will be a serious problem soon after integration," and Item 21, "I would not mind 'double dating' with a couple both of whom were of the other race") and dancing (Item 12, "Different racial groups mixing at school functions [dances, parties, etc.] will not be wise—it will only result in fights and ill feeling between races") a majority of the students expressed negative attitudes.

In many items, on the other hand, students in all classifications expressed overwhelmingly positive attitudes. For example, in Item 1, "If another race was integrated into my school, I would do my best to accept them as classmates and equals," over 90 per cent of the respondents expressed agreement. Likewise, in questions pertaining to integration in classrooms, athletics, and even fraternities, a large percentage of all classifications expressed favorable integration attitudes. In 22 out of the 29 "IA" items, a majority of all classifications expressed favorable integration attitudes, while in most of the remaining 7 items the several classifications were divided in attitudes, in general, the higher classifications as indicated above having the less favorable attitudes.

An interesting result of this study also is that the pattern is almost identical to that shown by the high school group in our earlier study (2). The high school group have considerably higher authoritarian attitudes than the college group and somewhat higher scores on the IA scale, but the clear pattern of positive attitudes with the decreasing positiveness in the close personal-social areas is present for both groups.

Conclusions

1. There is apparently little direct relationship between authoritarianism and negative integration attitudes for this population. The integration attitudes in the South are apparently so much a part of the culture that they are separated from the more generalized attitudes covered by authoritarianism. This hypothesis is lent additional weight by the fact that only in the graduate group some correlation was found. This, of course, is the group that through maturity and education may have gained insight into the relationship of the attitudes.

2. In general, it may be concluded that there is a generally positive attitude toward integration on the part of the white college student in West Texas, thus easing the widely expressed fear of serious trouble when integration comes. This conclusion is based on two facts. First, that the IA Scale means for all classifications lie well below a potential middle score of 116. This fact indicates that far more than 50% of the responses were those positive toward integration. This fact alone, however, would not be sufficient to justify this conclusion. The additional factor of the actual numbers of items to which an overwhelmingly positive response, as opposed to negative responses, was made adds substantial weight to the above statement. It should be noted that in 22 out of the 29 IA items the positive integration attitudes outnumber the negative ones in all five classifications, while in most of the remaining 7 the classes are divided in their response. Further, in many items there is better than a two to one ratio of positive to negative response.

3. Problems may be looked for in the areas of close personal-social contact such as socials, parties, dances and the like.

4. There is apparently a discrepancy between student and parental attitudes. Thus, where about 90 per cent of the students say they would do their best to accept members of the other race as classmates and equals and

the majority would accept them into clubs and even as good friends (Item 15), a definite majority would not bring these classmates and even good friends home with them due to fear of parental disapproval (Item 8).

Received April 8, 1957.

References

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. Greenberg, H., Chase, A. L., & Cannon, T. M., Jr. Attitudes of white and Negro high school students in a west Texas town toward school integration. *J. appl. Psychol.*, 1957, **41**, 27–31.

Journal of Applied Psychology
Vol. 42, No. 5, 1958

Anxiety Level and Score on a Biographical Inventory [1]

Donald H. Kausler and E. Philip Trapp

University of Arkansas

Siegel (5, 6) has recently developed a biographical inventory, The Biographical Inventory for Students (BIS), which shows promise as a predictor test, particularly in the area of academic achievement. He found, for example, that four of the 10 subscales, Action, Heterosexual Activities, Political Activities, and Dependence upon the Home, correlated significantly with college grade point average.

The relationship between scores on the Dependence upon the Home subscale (Dep) and grade point was curvilinear, with the higher grades corresponding to the extreme positions on the Dep continuum. In a follow-up study, Kausler and Little (4), relating Dep scores with grades in psychology courses, obtained a similar curve.

Siegel, in discussing the meaning of this curvilinear relationship, suggests that the *Ss* in the middle of the Dep range are more insecure than the *Ss* located at the two extremes. This difference in security accounts primarily for the difference reflected in academic performance.

The present authors prefer to interpret Siegel's findings in terms of a more clearly defined motivational variable. The curvilinear relationship suggests the operation of varying levels of anxiety drive at varying levels of dependency. In other words, the Dep subscale may, in effect, be measuring a motivational source of individual differences.

This study is directed at investigating the suggested relationship between the Dep subscale and anxiety level. It is hypothesized that: (*a*) the *Ss* scoring at the two extremes on the Dep subscale of the BIS will differ significantly in anxiety level from *Ss* scoring in the intermediate range; and (*b*) the *Ss* scoring at the two extremes on the Dep subscale will reflect lower anxiety level than the *Ss* scoring in the intermediate range. The hypothesized direction of the differences between the two groups follows Farber's suggestion (1) that performance level is apparently inversely related to anxiety drive level on complex learning problems.

Method

Subjects. Since only the BIS form for men is currently available, the *Ss* were 64 male students, predominantly freshmen and sophomores, enrolled in general psychology at the University of Arkansas.

Procedure. The entire BIS was administered with standard instructions to the *Ss* during a regular class period, though only the Dep subscale was used in this study. The *Ss* were informed that the objective of the inventory was to compare biographical information of students at the University of Arkansas with students in other universities.

Approximately one month later, the Taylor Manifest Anxiety Scale (7) was administered to the *Ss*, again during a regular class period. The scale, titled Biographical Inventory, was described to the *Ss* as an additional source of biographical information for comparing students at Arkansas with students elsewhere. The standard scoring procedures were employed.

Results

The *Ss* were divided into two groups on the basis of their scores on the Dep subscale. The scores ranged from -2 to $+20$, with a median of 9. The *Ss* in the first and fourth quartiles formed the extreme group ($N = 32$), and the *Ss* in the second and third quartiles formed the intermediate group ($N = 32$). The *Ss* were further classified according to their level of "manifest anxiety," applying the method usually followed in studies with the Taylor scale. The upper 25%, based on the distribution of scores for all 64 *Ss*, constituted a high "manifest anxiety" group (HA, $N = 16$), the lower 25% a low "manifest anxiety" group (LA, $N = 16$), and the middle 50% a middle "manifest anxiety" group (MA, $N = 32$).

Since both the Taylor scale and the BIS are ordinal scales, the hypotheses investigated were tested by a nonparametric test, χ^2. A

[1] The authors wish to thank Laurence Siegel and Educational Testing Service for their permission to use the BIS in this study.

Table 1

Dep Score and Anxiety Level: Observed and Expected Frequencies

	Extreme Dep	Intermediate Dep
HA	8 (2)	8 (14)
MA	16 (21)	16 (11)
LA	8 (9)	8 (7)

Note.—$\chi^2 = 12.36$ ($p < .01$).

3 × 2 contingency table was established comprising two groups, Extreme Dep (high and low scores on Dep) and Intermediate Dep (intermediate scores on Dep), and three classifications, High Anxiety (HA), Middle Anxiety (MA), and Low Anxiety (LA). The results of this analysis are presented in Table 1 below.

An examination of Table 1 reveals a large difference in "manifest anxiety" between the Extreme Dep and Intermediate Dep Groups and in the hypothesized direction. The Intermediate Group contains a preponderance of HA *Ss*, while the Extreme Group contains a high proportion of MA *Ss*. The overall χ^2 is 12.36, which is significant at the .01 level of confidence.

Discussion

Successful performance on complex tasks, such as college courses, is dependent upon interactions among a multiplicity of aptitude and motivational factors. The efficacy of biographical information in predicting performance on such tasks, therefore, depends upon the extent to which this information measures one or more of these factors. The biographical information derived from such scales as the BIS is interpreted by the authors as tapping primarily the motivational area. That is, certain facets of the individual's motivational system, depending, of course, on the content of the biographical data, express themselves through his personal history. The present study substantiates this interpretation in demonstrating a relationship between scores on the Dep subscale of the BIS and the motivational variable of anxiety as measured by the Taylor scale. However, a more rigorous interpretation of this relationship in terms of drive level is restricted by recent criticisms (2, 3) of the Taylor scale as a measure of drive.

The results of the present study also appear consistent with common clinical evidence on manifest anxiety. An individual's anxiety, ordinarily, should be more controlled or reduced when competition between his dependency and independency needs is low rather than high. Individuals scoring at the extreme positions on the Dep subscale are likely to experience relatively low competition between their dependency and independency needs because of marked subjugation or suppression of one need by the other. On the other hand, individuals scoring in the middle range, reflecting a closer balance of strength between their dependency and independency needs, are more likely to experience greater interneed conflict. The outcome would be evident in higher manifest anxiety.

Summary

The following two hypotheses were tested in the present study: (*a*) the *Ss* scoring at the two extremes on the Dep subscale of the BIS will differ significantly in anxiety level from *Ss* scoring in the intermediate range; and (*b*) *Ss* scoring in the intermediate range will reflect higher anxiety level. Anxiety level was measured by the Taylor Manifest Anxiety Scale. Sixty-four male students at the University of Arkansas served as *Ss*. In the analysis of results, a 3 × 2 contingency table was constructed comprising two groups, Extreme Dep and Intermediate Dep, and three classifications, High Anxiety, Middle Anxiety, and Low Anxiety. The over-all χ^2 was 12.36, significant at the .01 level of confidence, and in the hypothesized direction. The findings support the interpretation that the Dep subscale of the BIS is measuring the motivational variable of anxiety.

Received November 4, 1957.

References

1. Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.
2. Hill, Winfred F. Comments on Taylor's "Drive theory and manifest anxiety." *Psychol. Bull.*, 1957, **54**, 490–493.
3. Jessor, R., & Hammond, K. R. Construct Validity and the Taylor Anxiety Scale. *Psychol. Bull.*, 1957, **54**, 161–170.
4. Kausler, D. H., & Little, N. D. The BIS dependency scale and grades in psychology courses. *J. counsel. Psychol.*, 1957, **4**, 223–224.
5. Siegel, L. A biographical inventory for students: I. Construction and standardization of the instrument. *J. appl. Psychol.*, 1956, **40**, 5–10.
6. Siegel, L. A biographical inventory for students: II. Validation of the instrument. *J. appl. Psychol.*, 1956, **40**, 122–126.
7. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Journal of Applied Psychology
Vol. 42, No. 5, 1958

Development of a Criterion of Research Productivity[1]

Robert E. Stoltz

Southern Methodist University

The purpose of this study was to develop a criterion of research productivity for a physical science research organization where the usual, seemingly more objective criteria were not present or not applicable (i.e., number of publications).

Method

Subjects. The subjects in this study were all technically trained physical science research workers, holding at least a bachelor's degree in one of the physical science or engineering fields. The *Ss* were employed in research work covering a wide variety of science and engineering fields such as Mechanical Engineering, Reactor Metallurgy, Petroleum Chemistry, Ceramic Engineering, Nonferrous and Ferrous metallurgy. The raters were all research supervisors and technically trained persons. The raters' position in the company was analogous to the highest level of line supervision in the typical industrial organization. The *Ss* were all employees of a large midwestern research institute employing approximately 2,500 persons, of whom over one third might be considered to be technical research persons. The work of this institute is devoted largely to contract research; however, considerable time is given to pure, theoretical projects.

Development of the scale. Each of 27 research supervisors participated in a one and one-half hour interview with the investigator. This group of supervisors represented 57% of the supervisors at this level and constituted a cross section of all divisions of the institute. The interviews were conducted along a modification of Flanagan's Critical Incident technique (1). Informants were to select and describe three persons whom they considered to match the following descriptions: Most Productive, Most Nonproductive and Most Creative. The informants were then instructed to compare and contrast these persons, indicating to the investigator in what ways the persons they were describing behaved alike and in what ways they behaved differently in their research activities. The informants were instructed to use their own words, and to describe specific characteristics or events and to use whatever dimension for comparison or contrast they might wish.

At the conclusion of the series of interviews, the comments of the informants were examined and over 500 behavior statements were extracted. Eight psychologists were each asked to examine the items and to group them into whatever categories they felt the items represented. The investigator and two of the eight psychologists then reviewed the groupings that had been made, and finally decided on 15 categories into which the majority of the statements could be classed. Fifteen statements, or items, were selected from each category as most representative of the category by the three judges. Twenty-five other items, which could not be categorized with any agreement by the judges, but which were felt to be of interest and possible predictive value, were also chosen. While a discussion of the categories is necessary in order to establish the way in which the checklist was constructed, the categories themselves were not used in this inquiry and will not be referred to again. Additional information regarding the checklists may be found in an unpublished report by Stoltz (2). The items were reviewed again to determine whether or not they conformed to three requirements: (*a*) They must be observable by the persons who were to use the checklists; that is, administrative supervisors; (*b*) The behaviors mentioned in the items must be general enough to apply to persons in diverse fields of research inquiry and also general enough to apply to settings other than the present one. This was done so that in future research the same items might be used on new populations and make possible meaningful comparisons; (*c*) The general behavior mentioned in the item must have been suggested by a majority of the informants. All of the selected items met the above requirements or were reworded to conform to them.

Forty research supervisors and assistant research supervisors described 80 research personnel using this checklist (entitled the Productive Behavior Checklist). It should be pointed out that there is some possible contamination of the results as it was found necessary to use 15 of the original informants as raters in this phase of the study. Each rater was asked to describe two specific individuals: the most productive and the least productive man they knew. Twenty of the raters were instructed to rate the productive man first, the remaining 20 raters were asked to rate the least productive man first. Each item was to be rated on a five-point scale, with five indicating the highest possible rating, according to how well the item described the person being rated. The pages in the checklists were assembled in random order so that any tendency to rate items differentially due to serial position in the checklist would tend to be randomly distributed.

The checklists were then collected and the ratings for each item were converted into two indices. A discrimination index (DI) was computed for each

[1] The investigator is indebted to the staff of the Battelle Memorial Institute for their cooperation and assistance in conducting this study.

item by determining the amount of overlap between ratings for the productive and the nonproductive persons. On a sample of 50 randomly selected items, these DI's correlated .91 with *t* tests of the significance of the difference between the means of the two groups. A preference index (PI) was determined for each item by computing the mean rating given the item for both the productive and the nonproductive group. Two forced-choice type rating scales were constructed to use these items and their preference and discrimination indices. These scales were titled Research Behavior Description I and II (RBD I and RBD II). The items within each scale were composed of tetrads of items from the checklist, with each tetrad having two items with discrimination indices approximately 20 points greater than the discrimination indices of the remaining two, and with all items in the tetrad having nearly identical preference indices. RBD I contained 25 tetrads and RBD II contained 26 tetrads.[2]

Administration of the scale. Twenty-six divisions of the organization were selected to furnish raters for the study. These divisions constitute the principal units of organization of the institute. A division is to a great extent a functionally independent unit. A division comes into existence to handle a particular type of research problem and remains a division so long as that area of research remains important. If research activities in the area become increased, or if a particular aspect of the original problem becomes important enough to demand its own organization, the division enlarges or subdivides. Divisions range in size from 12 to 15 technical persons to over 100 technical persons. The division titles reflect the activity of the division. For example, some of the divisions covered in this study were Industrial Economics, Physical Chemistry, Ceramics, Nonferrous Physical Metallurgy, Light Metals, and Physical Metallurgy. Each division was sent six copies of RBD I and six copies of RBD II. Each division was also given a list of three men who were to be rated. These men were selected by the members of the personnel department on a random basis. The division supervisor was asked to select two supervisors to rate each man using both RBD I and RBD II. Following each rating the supervisor doing the rating was to indicate on each scale his over-all evaluation of the man being rated on a five-point scale, where five was the highest rating possible.

At the end of a three-week period, 139 RBD I and 136 RBD II forms had been returned out of the 156 copies of each form that had been distributed. It had previously been decided to use for analysis only those forms returned by this deadline. A score for each person rated was determined for each form by counting the number of times a rater had selected a higher discriminating item from each tetrad. An average RBD I rating was determined for each person rated by averaging the RBD I scores from his two raters. An average RBD II rating was determined for each person in the same manner. An over-all evaluation was determined for each person rated by averaging the rating given him on the five-point scale by each of the two raters. Since in some cases there was only data from one rater the *N*s in the accompanying table will not be identical. In addition, the over-all ratings were quite negatively skewed and were consequently converted to normalized *T* scores. Intercorrelations between the possible combinations were computed and are shown in Table 1.

[2] Copies of the checklists or RBD I and II may be obtained from the author.

Table 1

Research Behavior Description Reliabilities and Validities

Coefficient	Scale	*N*	*r*	*p*
Interrater Reliability	RBD I	62	.31	>.05
	RBD II	59	.62	>.01
	RBD I avg. vs. RBD II avg.	56	.74	>.01
	Five-Point Rating Criterion	65	.70	>.01
Validity	RBD I	139	.48	>.01
	RBD II	136	.60	>.01
	RBD I Average	62	.55	>.01
	RBD II Average	59	.74	>.01

Note.—RBD averages were correlated with person's average rating on five-point criterion scale to determine validity.

Results

Interrater reliability for RBD I was found to be .31. Interrater reliability for RBD II was found to be .62 and the reliability of the over-all rating on the five-point scale, after normalizing, was found to be .70. In view of the widely different reliability coefficient for the two forms an internal consistency item analysis was conducted to determine to what extent tetrad content influenced the coefficients. Although the tetrads had been randomly assigned to the two forms it was found that when biserial correlations were computed between tetrad scores and total scale scores, nine tetrads on RBD I and six tetrads on RBD II were found to be insignificantly related to total scale score. Moreover, four of the nine insignificant tetrads on RBD I had negative signs, while only one of the six insignificant tetrads on RBD II had negative signs. It would appear that the low reliability of RBD I is partially accounted for by its containing more poorly constructed tetrads. This view

is supported by the fact that when the biserial correlations were computed between the tetrads on RBD I and the over-all five-point rating, essentially the same nine items proved to be insignificantly related to the over-all rating.

Further evidence of the reliability of the scale was gained by determining the product moment correlation between the average RBD I and average RBD II ratings on each person for whom there are four ratings. This correlation was found to be .74.

The validity of the scales was assessed in several ways. First, each person's rating was correlated with his over-all rating on each form. RBD I was found to have a validity coefficient of .48 and RBD II to have a coefficient of .60. When the person's average rating on each form was correlated with his average over-all rating on the five-point scale, RBD I showed a coefficient of .55 while RBD II showed a correlation of .74.

Discussion

RBD II showed quite satisfactory interrater reliability and validity. RBD I did not show adequate interrater reliability, but this was no doubt due to the fact that this form contained a much greater number of insignificant, no doubt poorly constructed, tetrads than did RBD II.

Validity coefficient based on the average ratings of two raters showed increased validity over either scale alone as one might expect. The correlation between the average ratings on the two scales indicates that even with the unreliability of RBD I, the two scales are measuring much the same thing.

It appears that it is possible to develop a forced-choice rating scale to determine the extent of a person's productive research behavior, at least in so far as it is evaluated by his superior. The extent to which ratings on the RBD scales correlate with other measures of research productivity is not yet known. Neither is it known to what extent subordinates' evaluation of productivity might agree with superiors' evaluations. All of these problems are yet to be answered by future applications of the scales and their derivatives.

Perhaps the most interesting fact of the entire study is that the rating scales predicted as well as they did when one considers the heterogeneous character of the divisions used in the study. It would appear to indicate that if the rating forms are evaluating productive behavior, the same type of behavior is given merit in one division as in another even though the activities of the divisions may be quiet different. There appear to be two possible, perhaps related, explanations. One, that a person's productivity is influenced most by characteristic work habits or attitudes that the person shows in all challenging activities and is somewhat independent of job knowledge. Or two, that supervisors are all responding to a stereotype of the productive person that they reward in each division and which may have nothing to do with over-all productivity in a more objective sense. The present study does not indicate which of these two interpretations is correct. One would prefer the former explanation if less concern is in developing and understanding the productive process; but if our concern is predicting the ability of a person to rise in company esteem and status, the latter is of just as much benefit.

Summary

A forced-choice rating scale designed to determine the extent of a person's productive research behavior was developed at a Midwestern physical science foundation. Of the two experimental scales developed the better form showed an interrater reliability coefficient of .62 and a validity of .60. When the ratings of two raters were averaged the validity of the scale increased to .74.

Received December 16, 1957.

References

1. Flanagan, J. C. The critical incident technique. *Psych. Bull.*, 1954, **51**, 327–358.
2. Stoltz, R. E. A study of productivity in a research setting. Unpublished doctoral dissertation, Ohio State Univer., 1956.

Journal of Applied Psychology
Vol. 42, No. 5, 1958

Performance as a Function of Control-Display Relations, Positions of the Operator, and Locations of the Control [1]

Michael Humphries [2]
University of Toronto, Canada

The ease with which a particular combination of a control and a display can be learned and operated is frequently determined by the nature of the control-display movement relation employed. Operators appear to expect, on the basis of prior experience, that certain control movements should produce certain specific display changes (15). If these movement relationships are not present, performance is frequently impaired. These effects have been demonstrated on the Two-Hand coordinator (4, 17, 23, 25, 29, 32), and Turret Pursuit Apparatus (17, 21, 22, 33), the Mashburn Apparatus (1, 2, 10, 17, 18, 19, 20, 26), the Toronto Complex Coordinator (15, 31), and for a variety of levers (6, 11, 13, 19, 24), rudder pedals (13, 14, 24) and knobs (7, 8, 12). The basic conclusion from these experiments is that the movement of the control should be in the same plane and in the same direction as the resulting display movement in order to obtain the highest performance level and the fastest learning (5, 16).

Although the directions and planes of movement of the two elements should be consistent in most cases, there are a few situations in which this generalization does not apply. In systems employing cranks or knobs it is difficult to define consistency since rotary movement of a crank or knob is converted into linear movement on the display. This is the case on most two-hand coordinators. Also, in some simple tracking tasks using levers and where the operator does not work under stress conditions, the principle of consistency does not apply; controls can be connected to the display so that the movement of the control is inconsistent with display movement without impairing performance level (9, 27, 28, 34). In view of these limitations the remainder of the present work should be considered in terms of the situations where only relatively complex tasks are used and where linear control movement results in linear display change.

It has also been demonstrated that the position of the operator relative to the display is an important factor in determining the level of performance. In an experiment soon to be reported, the operator was seated to one side of the display so that his frontal plane was at right angles to the frontal plane of the display. Under these conditions, with the control in front of him the operator had to turn his head slightly to the left in order to see the display. It was found that this body position greatly improved the level of performance obtained from a control-display relationship which normally leads to an extremely low level of performance when the *S* is seated directly in front of the display (15). This suggested that there might be an interaction between the effects of body orientation and control-display relations. Since there was neither a theory nor sufficient empirical evidence to suggest what the hypothesized interaction might be, a model was constructed for this purpose.

The model was designed to predict the effect on performance of variations in control position, body orientation and control-display relations for the Toronto Complex Coordinator (31). Since the model was not completely successful and as it is not possible to describe it in a brief paper, only the empirical findings will be reported at this time.

Method

Apparatus

The apparatus used was the Toronto Complex Coordinator (TCC). This unit has been described

[1] This research was conducted at the University of Toronto under Research Grant No. 9401-02 from the Defence Research Board of Canada.

This paper represents part of a dissertation submitted to the University of Toronto in 1957 in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to acknowledge the assistance of A. H. Shephard of the University of Toronto.

[2] Now with the Defence Research Medical Laboratories, Toronto, Canada.

in the literature (**15, 31**); consequently only a brief description will be given in this paper.

The *S* is seated before a vertical panel of 81 light assemblies, each consisting of an inner green disc and an outer red ring. These lights are covered by an opaque glass screen. Only one green disc and one red ring can be seen at a time. Position of the ring is determined by a stepping switch; that of the green disc by the position of the *S*'s control. The *S* is required to move an airplane-type control stick so that the green disc within the ring is illuminated. When this has been accomplished, a new ring, or target, appears elsewhere on the screen. The task in this case is discrete, self-paced tracking.

The measure of performance used was the number of matches made in a five-minute practice period. A match is recorded each time the *S* places the green disc within the illuminated target.

Experimental Variables

Three classes of independent variables were treated in the present work: control position, body orientation, and control-display relationships. These three factors are discussed below.

1. Control Positions: The majority of research on the TCC has been done under conditions where the control stick is perpendicular to the floor. This means that there can be no consistency between control movement and vertical changes on the display; the operator can move the control in the horizontal and normal directions, but not in the vertical. For example, the "expected" or "standard" control-display relationship has required the operator to push the stick away from himself to make the green light move up the display (**15, 31**).

In a number of experiments (**5**) it has been demonstrated that the planes of movement of the control and the display should be the same to produce the highest level of performance. Consequently it was thought that the best level of performance would be obtained on the TCC if the control-unit were arranged so that both the vertical and the horizontal movement relations of the control and the display were the same. For this reason the control unit was mounted on a stand so that the control stick projected towards the *S*, parallel to the floor and perpendicular to the frontal plane of the display. In this way the other variables to be discussed can be compared under two positions of the control, one of which permits partial consistency in the planes of movement, and the other which permits complete consistency.

2. Body Orientation: As mentioned previously, body reorientation made a usually difficult task very much easier than it was when the operator sat directly in front of the display. Since the orientation of the *S* could differentially affect the various control-display relationships to be discussed later, it was decided that this factor should be systematically varied. The three body positions tested are illustrated in Fig. 1. These positions have been designated according to the clock-face numerals, 3, 6, and 9.

3. Control-Display Relationships: The nature of the electrical connection between the control and the display of the TCC makes it possible to arrange a large number of control-display relationships. Any given arrangement, or task, can be defined by specifying the relationship between control movement and display change, where these directions are defined in terms of the display itself, regardless of the position of the control unit or of the operator. Although it is possible to arrange a number of tasks, only four are to be used in the present work. To be consistent with previously employed terminology, these four tasks have been defined in Table 1. In the column of spatial relationships, the first letter of each pair refers to the direction of control movement relative to the display, while the second in each pair refers to the direction of movement of the green light on the display. The letters A, T, R, L, U, and D refer to the directions, away, towards, right, left, up, and down, respectively. For example, in Task A, when the stick is moved away from the display (A) the green light moves down the panel (D).

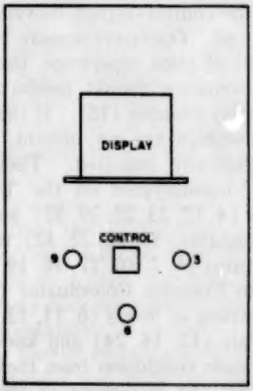

FIG. 1. Plan view showing the position of the control, display, and three positions of the operator.

Table 1

Control-Display Relationships for Four Tasks on the TCC with the Control-Stick Mounted in the Vertical Position

Task	Spatial Relations			
A	A–D	T–U	R–R	L–L
B	A–U	T–D	R–L	L–R
C	A–R	T–L	R–U	L–D
D	A–L	T–R	R–D	L–U

Table 2

Control-Display Relationships for Four Tasks on the TCC with the Control-Stick Mounted in the Horizontal Position

Task	Spatial Relations			
A′	D–D	U–U	R–R	L–L
B′	D–U	U–D	R–L	L–R
C′	D–R	U–L	R–U	L–D
D′	D–L	U–R	R–D	L–U

The above definitions of the four tasks do not apply when the control-unit is mounted on a stand due to the difference in vertical movement relations. In recognition of this fact, the four tasks with the control-unit on the stand are defined in Table 2. The meaning of the letters is the same as specified above.

Experimental Design

The experimental design was a $4 \times 3 \times 2$ factorial (four tasks, three body positions, and two control positions) with 24 independent cells, each cell consisting of 10 independent observations. This design permitted 23 orthogonal comparisons to be made.

Procedure

Each *S* was given a standard set of verbal instructions. The control-display relation was demonstrated and the *S* was shown how to make a match. All *S*s were instructed to make as many matches as possible. Each *S* worked continuously for five minutes, on only one of the experimental conditions.

Results

The mean number of matches made by each group is shown in Fig. 2. This figure shows how performance on each task varies with the position of the operator.

From these curves a number of trends are apparent:

1. In Position 6: the levels of performance on Tasks C, D, C′, and D′ are similar. The levels of performance resulting from practice on Tasks B and B′ are slightly two-thirds the distance between the levels found for Tasks A and A′ on the one hand, and the levels for C, D, C′, and D′ on the other. Tasks B and B′ produce the same levels of performance; so do Tasks A and A′. It is concluded, therefore, that the positions of the control used in this study do not influence performance on these tasks when the operator is seated in Position 6.

2. In Positions 3 and 9: the levels of performance on Tasks A and B are lower in Positions 3 and 9 than they are in Position 6, although the differences between the levels remain the same. It is concluded that the best performance on these tasks occurs when the operator sits facing the display. This is not the case for Tasks C and D, however. The performance levels for these tasks are increased by rotation of the operator, but the amount of improvement in each case depends on the direction of rotation. In Position 9, Task C appears to be easier than D, while in Position 3 the opposite is true.

3. There is no apparent effect on performance produced by rotation of the operator when he works on Tasks A′, B′, C′, or D′, since these tasks have essentially the same level of performance in Positions 3 and 9 that they have in Position 6.

Statistical analysis of the data illustrated in Fig. 2 supports the above generalizations. A summary of the analysis of variance is presented in Table 3.

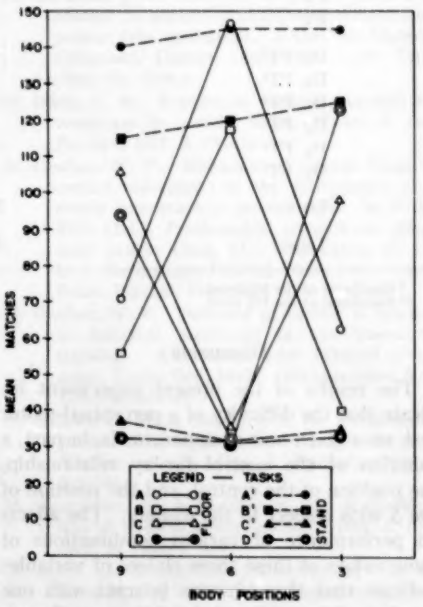

FIG. 2. Performance as a function of control-display relations, positions of the operator and locations of the control.

Table 3

Summary Analysis of Variance Table of the Effects of Four Tasks, Three Positions of the Operator and Two Positions of the Control

Source	*df*	Mean Square	*F*
Body Position (D)			
DL (Linear)	1	4.2250	—
DQ (Quadratic)	1	39.6750	—
Stand vs. Floor (P)	1	47.7042	—
Tasks (T)			
T′ ((A+A′)−(B+B′))	1	13,996.8000	34.53**
T^2 ((C+C′)−(D+D′))	1	91.8750	—
T^3 ((A+A′)−(B+B′))− ((C+C′)−(D+D′))	1	148,553.5042	366.49**
DL P	1	354.0251	—
DL T′	1	12.0125	—
DL T^2	1	2,387.1125	5.89*
DL T^3	1	518.4000	—
DQ P	1	69.0083	—
Dq T′	1	397.8375	—
Dq T^2	1	57.0375	—
Dq T^3	1	73,112.0333	180.37**
PT′	1	22.5333	—
PT^2	1	185.0083	—
PT^3	1	143,521.5042	354.07**
D_L PT′	1	255.6125	—
D_L PT^2	1	1,320.3125	3.26
D_L PT^3	1	3,240.0000	7.99**
D_q PT′	1	87.6042	—
D_q PT^2	1	250.1042	—
D_q PT^3	1	66,646.5333	164.42**
Error	216	405.3458	
Total	239		

* Significant at the 5% level.
** Significant at the 1% level.

Discussion

The results of the present experiment indicate that the difficulty of a perceptual-motor task on a given unit of apparatus is, in part, a function of the control-display relationship, the position of the control, and the position of the *S* with respect to the display. The effects on performance of various combinations of some values of these three classes of variables indicate that these factors interact with one another. It was observed that some tasks are made more difficult while others are made easier by rotation of the *S* away from a position directly in front of the display. This indicates that knowledge of the control-display relationship alone is not sufficient to predict the relative levels of difficulty.

Also, it is not possible to predict the effects of body rotation on the various tasks without considering the position of the control.

These observations have serious implications for equipment design. Under conditions where the *S* regards the display over his shoulder, and perhaps also at other angles of regard, the best level of performance and the greatest facility in learning can be achieved

by using somewhat unconventional control-display relationships. This conclusion probably holds only for certain types and positions of controls and displays. Consequently, until others have been tested the present results should not be used as the basis for general design principles.

Summary

Previous research has shown that control-display relations, position of the control and position of the operator are important factors determining the level of performance. Since these variables have usually been studied in isolation, an experiment was designed to investigate the interaction between them.

Twenty-four groups of male *Ss* practiced for five minutes on the Toronto Complex Coordinator. Each group worked on only one of the 24 combinations of experimental conditions. The results indicate that for the same apparatus, knowledge of the control-display relations alone is not sufficient to predict the relative levels of performance.

Received December 30, 1957.

References

1. Adams, J. A. The evaluation of "difficulty of task" under different conditions of performance on the modified Mashburn Apparatus. Office of Naval Research, Special Devices Center, Port Washington, L. I., N. Y., 1949. (Tech. Rep. SDC 57-2-8.)
2. Adams, J. A. The influence of the time interval after interpolated activity on psychomotor performance. Air Training Command, Hum. Resour. Res. Cent., *Res. Bull. 55-11*. San Antonio, Texas: Lackland AFB, 1952.
3. Andreas, B. G. Bibliography of perceptual-motor performance under varied display-control relationships. Rochester, N. Y.: Univer. Rochester, 1953. (Contract AF 30(602)-200 Sci. Rep. No. 1.)
4. Andreas, B. G., Green, R. F., & Spragg, S. D. S. Transfer effects between performance on a following tracking task (modified SAM Two-Hand Coordinator Test) and a compensatory task (modified SAM Two-Hand Pursuit Test). *J. Psychol.*, 1954, **37**, 173–183.
5. Andreas, B. G., & Weiss, Bernard. Review of research on perceptual-motor performance under varied display-control relationships. Rochester, N. Y.: Univer. Rochester 1954. (Contract A.F. 30(602)-200, Sci. Rep. No. 2.)
6. Bilodeau, E. A. Modifications of directions of movement preference with independent variation of two stimulus dimensions. Air Training Command, Hum. Resour. Res. Cent., *Res. Bull., 51-12*. San Antonio, Texas: Lackland AFB, 1951.
7. Carter, L. F., & Murray, N. L. A study of the most effective relationships between selected control and indicator movements. In P. M. Fitts (Ed.), *Psychological research on equipment design*, Chap. 10. Washington, D. C.: U. S. Government Printing Office, 1947. (AAF Aviat. Psychol. Prog. Res. Rep. No. 19.)
8. Fitts, P. M., & Simon, C. W. Some relations between stimulus patterns and performance in a continuous dual-pursuit task. *J. exp. Psychol.*, 1952, **43**, 428–436.
9. Foxboro Company. A study of factors determining accuracy of tracking by means of hand-wheel control. Report to the Services No. 71. Washington, D. C.: Division 7, National Defence Research Committee, 1942. (Office of Scientific Res. and Develpm. Rep. No. 3451.)
10. Gagne, R. M., Baker, Katherine E., & Wylie, Ruth C. The effects of an interfering task on the learning of a complex motor skill. Port Washington, L. I., N. Y.: Office of Naval Research, Special Devices Cent., 1949. (Tech. Rep. SDC 316-1-9.)
11. Gardener, J. F. Direction of pointer motion in relation to movement of flight-controls—cross-pointer type instrument. USAF, Air Materiel Command, Dayton, Ohio, 1950. (AF Tech. Rep. No. 6016.)
12. Gibbs, C. B. Transfer of training and skill assumptions in tracking tasks. *Quart. J. exp. Psychol.*, 1951, **3**, 99–110.
13. Grether, W. F. Efficiency of several types of control movements in the performance of a simple compensatory pursuit task. In P. M. Fitts (Ed.), *Psychological research on equipment design*, Chap. 17. Washington, D. C.: U. S. Government Printing Office, 1947. (AAF Aviat. Psychol. Prog. Res. Rep. No. 19.)
14. Grether, W. F. Direction of control in relation to indicator movement in one-dimensional tracking. AAF Hdqrs, Air Materiel Command, Engin. Div., 1947. (Memorandum Rep. TSEAA—694-4G.)
15. Humphries, M., & Shephard, A. H. Performance on several control-display arrangements as a function of age. *Can. J. Psychol.*, 1955, **9**, 231–238.
16. Humphries, M., & Fletcher, N. C. A classification of tasks in motor learning. *Can. J. Psychol.*, in press.
17. Lewis, D., & McAllister, Dorothy, E. An investigation of individual susceptibility to interference. Office of Naval Res., Special Devices Cent., Port Washington, L. I., N. Y., 1950. (Tech. Rep. SDC 938-1-10.)

18. Lewis, D., McAllister, Dorothy E., & Adams, J. A. Facilitation and interference in performance on the modified Mashburn Apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, **41**, 247–260.
19. Lewis, D., Shephard, A. H., & Adams, J. A. Evidences of association interference in psychomotor performance. *Science*, 1949, **110**, 271–273.
20. Lewis, D., & Shephard, A. H. Devices for studying associative interference in psychomotor performance: I. The modified Mashburn Apparatus. *J. Psychol.*, 1950, **29**, 35–46.
21. Lewis, D., & Shephard, A. H. Devices for studying associative interference in psychomotor performance: IV. The Turret Pursuit Apparatus. *J. Psychol.*, 1950, **29**, 173–182.
22. Lewis, D., & Shephard, A. H. Prior learning as a factor in shaping performance curves. *Proc. nat. Acad. Sci.*, 1951, **37**, 124–131.
23. Lewis, D., & Smith, P. N. Retroactive facilitation and interference in performance on the modified Two-Hand Coordinator. Office of Naval Res., Special Devices Cent., Port Washington, L. I., N. Y., 1951. (Tech. Rep. SDC 166-00-2.)
24. Loucks, R. B. An experimental study of the effectiveness with which novices can interpret a localizer-glidepath approach indicator. United States Air Force, Air Materiel Command, Dayton, Ohio. USAF Technical Report No. 5959, 1949.
25. McAllister, Dorothy, E. Retroactive facilitation and interference as a function of level of learning. *Amer. J. Psychol.*, 1952, **65**, 218–232.
26. McAllister, Dorothy E., & Lewis, D. Single-trial-per-task versus multiple-trial-per-task in the acquisition of skill in performing several similar tasks. *Proc. Iowa Acad. Sci.*, 1950, **57**, 407–416.
27. Mitchell, M. J. H. Direction of movement of machine controls: III. A two handed task in a discontinuous operation. Cambridge, England: Psychol. Lab., Ministry of Supply, 1947. (SM 10018 (S).)
28. Mitchell, M. J. H. Direction of movement of machine controls: IV. Right- or left-handed performance in a continuous task. Cambridge, England: Medical Research Council, Unit of Applied Psychology, 1948. (MRC 48/371, APU No. 85.)
29. Norris, Eugenia B., & Spragg, S. D. S. Performance on a following tracking task (modified SAM Two-Hand Coordinator Test) as a function of the planes of operation of the controls. *J. Psychol.*, 1953, **35**, 107–117.
30. Shephard, A. H. Losses of skill in performing the standard Mashburn task arising from different levels of learning on the reversed task. Office of Naval Res., Special Devices Cent., Port Washington, L. I., N. Y., 1950. (Tech. Rep. SDC 938-1-9.)
31. Shephard, A. H. The Toronto Complex Coordinator. *Canad. J. Psychol.*, 1955, **9**, 7–14.
32. Shephard, A. H., & Lewis, D. Devices for studying associative interference in psychomotor performance: II. The modified Two-Hand Coordinator. *J. Psychol.*, 1950, **29**, 53–66.
33. Speeth, W., & Lewis, D. The effects of alternating practice on the performance of two antagonistic motor tasks. Office of Naval Res., Special Devices Res. Cent., Port Washington, L. I., N. Y., 1950. (Tech. Rep. SDC, 938-1-6.)
34. Vince, M. A., & Mitchell, M. J. H. Direction of movement of machine control (continued). Cambridge, England: Ministry of Supply. Psychological Lab., 1946. (SM 2861 (S).)

Journal of Applied Psychology
Vol. 42, No. 5, 1958

The Postwar Occupational Adjustment of Emotionally Disturbed Soldiers

John B. Miner [1]

Personnel Research Division, The Atlantic Refining Company

and James K. Anderson

Conservation of Human Resources Project, Columbia University

Job applicants with a history of severe emotional disturbance have recently become a major problem for those concerned with personnel selection. Some employers have adopted an explicit policy forbidding the hiring of people known to have such a history on the assumption that they do not represent good employment risks. Probably the majority, however, have established no specific policy, preferring to leave the matter to the discretion of individual employment interviewers. Prior to World War II the problem of establishing policy was, in fact, rarely raised, since interviewers found it very difficult to obtain specific information from applicants.

Since the war it has become common practice to examine the discharge papers of applicants, or at least inquire about the conditions under which a man was separated from military service. Consequently, many personnel managers have become increasingly aware of the very considerable incidence of mental disease in the population. Unfortunately, however, little information is available as to the subsequent work histories of men who have suffered from severe emotional disturbances. Employers, caught between a very definite social pressure to hire all disabled veterans and their own misgivings as to the value of men with a history of neurosis or psychosis as workers, have lacked the factual data necessary to making valid policy decisions.

The present paper represents an attempt to, at least in part, fill this gap, utilizing data gathered in connection with the Conservation of Human Resources Project's study of ineffective performance during World War II. More specifically an effort is made to investigate some of the assumptions which, for lack of specific information, have on occasion guided personnel policy in this area. These assumptions which form the hypotheses of this study may be stated as follows:

1. Men with a history of neurosis or psychosis will reveal their difficulties in occupational adjustment by having a higher incidence of unemployment than other men from similar backgrounds.

2. Men with a history of neurosis or psychosis will reveal their difficulties in occupational adjustment by failing to obtain jobs at as high a level as other men from similar backgrounds.

Procedure

The *Ss* for the study were selected from a special sample consisting of five per cent of all enlisted men separated from the army during and immediately after World War II. Certain information on the men in the five per cent sample including date of induction and reason for separation had been punched on I.B.M. cards and these cards were retained at the Adjutant General's Office in Washington. From this large sample, all men who had been inducted during the last four months of 1942 and who had been prematurely discharged for either neurosis or psychosis were selected. Both the neurotics and the psychotics were then divided into six groups, using certain length and type of service categories. These categories were introduced in order to facilitate analyses beyond the scope of the present paper and have been described in detail elsewhere (3). The 12 groups (six neurotic and six psychotic) were reduced still further in size using a random selection process. For reasons extraneous to the present study, the number of cases included in each of the 12 smaller samples was not directly proportionate to the frequencies in the five per cent sample.

Since, for the purposes of the ensuing analyses, data on the two diagnostic groups (the neurotics and the psychotics) only were desired, a weighting

[1] This research was carried out while the senior author was on the staff of the Conservation of Human Resources Project, Columbia University.

process was used to combine the frequencies for the 12 length and type of service samples. The weighting factors applied to data from the six neurotic and six psychotic samples were derived from the actual frequencies present in the five per cent sample. Through their use it has been possible to obtain composite estimates of the figures that would have been obtained if it had been possible to study all enlisted men inducted into the army during the last four months of 1942 and later discharged prematurely with either a neurotic or psychotic condition. The diagnostic composition of these two samples has been indicated elsewhere (3).

In addition to the neurotic and psychotic samples, a control group was drawn from the original five per cent sample. This group consisted of enlisted men inducted in the last four months of 1942 who were subsequently honorably discharged from the army at the end of the war as a result of demobilization. The *Ss* included in the control group were selected so as to individually match every other man in the neurotic sample on four variables: education, race, age, and county of residence. It was not possible to obtain perfect matching on the latter variable, but every effort was made to keep deviations to a minimum. Thus, if a man from a rural county in Georgia could not be matched with another man from the same county, he was at least matched with a man from another similar county in his own or an adjacent state. The results of the matching on education, race, and age are presented in Tables 1, 2, and 3 respectively. These tables also contain distributions for the psychotic sample, which did not serve directly as a basis for the selection of the control group. Nevertheless, neither the psychotics nor the neurotics differ reliably from the controls on any of these three variables. Chi-square analysis using the .05 level as a criterion of adequacy indicates that the matching process was satisfactory.

Table 1

Percentage of Neurotics, Psychotics, and Controls Attaining Various Educational Levels Prior to Army Induction

		School Years Completed				
Group	*N*	0–6	7–8	9–11	12	13+
Neurotics	225	18.7	23.6	28.0	25.3	4.4
Psychotics	115	25.2	18.3	31.3	17.4	7.8
Controls	102	18.6	15.7	31.4	30.4	3.9

Table 2

Percentage of Neurotics, Psychotics, and Controls of White and Negro Race

		Race	
Group	*N*	White	Negro
Neurotics	225	85.3	14.7
Psychotics	115	80.0	20.0
Controls	102	85.3	14.7

Table 3

Percentage of Neurotics, Psychotics, and Controls of Various Ages at Army Induction

		Years				
Group	*N*	18–19	20–24	25–29	30–34	35+
Neurotics	225	7.6	55.6	17.3	10.2	9.3
Psychotics	115	9.5	58.3	17.4	8.7	6.1
Controls	102	8.8	55.9	22.5	9.8	3.0

The *N* values noted in these tables require some further explanation. It was not possible to obtain adequate follow-up information on the employment status of all *Ss*. Postwar information was available on 88.7 per cent of the 115 controls, 97.4 per cent of the 231 neurotics, and 95.8 per cent of the 120 psychotics. These smaller numbers have been employed in this and all ensuing analyses. Since the total neurotic sample rather than the half which served as a basis for matching has been utilized, the neurotics and controls do not appear as perfectly matched.

Extensive information on all *Ss* was obtained from the files of the Army Records Center in St. Louis. These files contain service records, medical records, and separation board proceedings. In addition, considerable follow-up data have been accumulated. The VA files were checked through 1953. Many of the *Ss* obtained disability compensation and medical treatment from the VA. Others have received extensive education and training, frequently coupled with vocational guidance interviews and testing. As a result of these contacts, the VA has built up a rather complete dossier on the postwar employment of these men. When the VA data were, for some reason, incomplete, a situation which was most likely to arise among the control *Ss*, they were supplemented through the use of questionnaires mailed to the veterans' homes. These questionnaires, which inquired specifically into a man's occupational and educational adjustment, were mailed during the latter part of 1954. It was only in those few instances when an *S*'s VA file contained no information on postwar employment and it was impossible to contact him by mail that a man had to be eliminated from the sample because of inadequate information.

For the purposes of the present analyses, only the most recent available data have been employed. A man was categorized as a student, unemployed, or working as of the last date such information was obtained, whether from the VA files or the questionnaire. These terminal dates are indicated in Table 4. Although the information on the controls tends to

Table 4

Year of Last Available Follow-up Information for Neurotics, Psychotics, and Controls

Year	Group		
	Neurotics (%)	Psychotics (%)	Controls (%)
1945–46	3.1	4.2	1.0
1947–48	4.1	7.0	5.9
1949–50	12.4	7.0	6.9
1951–52	26.2	20.9	8.8
1953–54	54.2	60.9	77.4

be somewhat more recent than that on the psychiatric groups, the controls were also discharged from the Army later. Thus, the period since leaving military service may be considered as of essentially the same duration for all groups.

In order to investigate the first hypothesis, estimates of the extent of unemployment among prematurely discharged neurotics and psychotics were obtained using the weighting process previously described. The control group figure was derived by applying the six weighting factors developed for use with the neurotics to the control data as well. The incidence of unemployment among the neurotics and psychotics was compared with that for the matched controls using chi-square techniques.

The investigation of the second hypothesis requires a specification of the occupational level attained by those men who were employed at the time of follow-up. For this purpose the system of classification according to intellectual demand described in *Intelligence in the United States* (2) was used. The system assigns each occupation to one of four levels, depending on the intelligence required for adequate performance. The levels may be generally described as follows:

Level 1: High level professional workers, executives, those involved in the most technical and complex clerical and sales work, and large scale farmers.

Level 2: Retail managers, the more highly skilled workers, skilled clerical workers, foremen, wholesale salesmen, technicians, lower level professional workers, and relatively large scale farmers.

Level 3: Lower level skilled workers, the semi-skilled, routine clerical workers, retail sales clerks, most farmers, and proprietors of relatively simple businesses.

Level 4: Unskilled workers.

After the occupational level of all employed members of the neurotic, psychotic, and control samples had been established and the distributions weighted,[2] the values for the neurotics and psychotics were compared with those for the controls. In addition, the number of neurotics and psychotics employed at each level was compared with similar figures obtained as a result of categorizing the occupations found in a representative sample of the employed population as of 1953. Detailed descriptions of this sample, which contained 745 *Ss*, have been presented elsewhere (2, 4).

A second analysis was also carried out using the total neurotic, psychotic, and control samples. In this instance the occupational level of the last known job held by those men who were either students or unemployed at the time of follow-up was determined and these occupational distributions added to those of the employed. Comparisons between groups were then made in the same manner as previously indicated.

Results

The incidence of unemployment in the representative sample of neurotics inducted in late 1942 is 13.4 per cent. This figure compares with a control value of 7.4 per cent. The difference is not reliable ($\chi^2 = 2.58$, $P > .10$). Among the psychotics, on the other hand, unemployment is quite prevalent with 48.9 per cent out of work at the time of follow-up. When compared with the controls, this figure yields a highly reliable difference ($\chi^2 = 44.77$, $P < .01$). Thus, the results obtained with the psychotic group lend considerable support to the first hypothesis. There is no clear-cut evidence, however, that the incidence of unemployment among former neurotics is any higher than among other men from similar backgrounds.

When the occupational distribution of the employed neurotics is compared with that of the controls (Table 5), a highly reliable difference is obtained ($\chi^2 = 13.51$, $P < .01$). Detailed analysis indicates that this results from an underconcentration of neurotics in the Level 2—skilled occupations—and a corresponding increase in Level 4—unskilled workers—(χ^2s = 6.76 and 6.71 respectively, $P < .01$ in both cases). When a similar analysis using the employed population as of

[2] As with the unemployment figure, weighting of the control distribution was carried out using the six factors developed for use with the *neurotics*. The use of the factors derived from data on the incidence of *psychosis* within the five per cent sample would not have brought about an appreciable change in the results.

Table 5

Percentage of the Employed Neurotics, Psychotics, and Controls and of the Employed Population Working at Each Occupational Level

Group	*N*	Occupational Level 1	2	3	4
Neurotics[a]	193	4.6	21.2	48.7	25.5
Psychotics[a]	56	7.6	30.6	40.5	21.3
Controls[a]	90	.6	35.6	51.8	12.0
Employed Population	745	12.2	31.7	38.4	17.7

[a] Based on weighted data.

1953 is carried out, this tendency to lower level employment among the neurotics is even more pronounced ($\chi^2 = 22.96$, $P < .01$). The neurotics are reliably below the population in Level 1 and Level 2 employment and reliably above at Levels 3 and 4 (χ^2s = 9.38, 8.12, 6.84, and 5.93, all with $P < .01$ except at Level 4 where P is $< .02$). The difference between the findings obtained with the matched controls and the employed population is attributable to the relatively greater incidence of Level 1 employment in the population as a whole and the decreased proportion of Level 3 workers.

Similar results are obtained when the last known occupation and the total neurotic sample form the basis for the analysis (Table 6). The neurotics and neurotic controls still differ ($\chi^2 = 14.38$, $P < .01$) and this difference is still attributable to the lower, Level 2 frequency and higher, Level 4 frequency among the neurotics (χ^2s = 6.13, $P < .02$ and 8.90, $P < .01$, respectively). The neurotic-employed population comparison is also reliable ($\chi^2 = 34.54$, $P < .01$), and again the underconcentration of neurotics in Levels 1 and 2 and the overconcentration at Level 4 occurs (χ^2s = 12.78, 10.68, and 16.93, respectively, all with $P < .01$). A reliable difference no longer appears at Level 3.

Table 6

Percentage of Neurotics, Psychotics, Controls, and Employed Population Working at Each Occupational Level at the Time of Last Known Employment

Group	*N*	Occupational Level 1	2	3	4
Neurotics[a]	225	3.9	20.4	45.3	30.4
Psychotics[a]	115	3.6	17.1	29.3	50.0
Controls[a]	102	.7	33.1	51.3	14.9
Employed Population	745	12.2	31.7	38.4	17.7

[a] Based on weighted data.

In spite of the consistent evidence of lower level employment among the neurotics, no such result is obtained when the distribution of the employed psychotics is compared with that of the controls (Table 5). The difference is not reliable ($\chi^2 = 5.80$, $P > .10$). Similarly, the psychotic-employed population comparison does not produce evidence of a reliable difference ($\chi^2 = 1.41$, $P > .10$).

When, however, the last known occupations of the students and unemployed are added to the employed distribution (Table 6), a significant chi-square value is obtained ($\chi^2 = 34.52$, $P < .01$). The psychotics are less likely to have been employed in Levels 2 and 3 occupations than the controls and much more likely to have been working at Level 4 jobs (χ^2s = 7.53, 11.29, and 29.99, all with $P < .01$). When the psychotics are compared with the employed population of 1953, a similar result is obtained ($\chi^2 = 63.19$, $P < .01$). The psychotics are underrepresented at Levels 1 and 2 and heavily concentrated at Level 4 (χ^2s = 7.56, 10.10, and 60.49, respectively, all with $P < .01$).

Discussion

Although the possibility that the lower level of employment among the neurotics and the high incidence of unemployment among the psychotics might be attributable to differences in education, race, or age has been ruled out through the use of matched controls, it is possible that these findings are, at least in part, attributable to a restriction of the occupational opportunities available to men who have received Certificate of Disability Discharges for emotional reasons. There is reason to believe, however, that this is not a major factor contributing to the results. Very rarely do statements appear in either the VA records or the questionnaires of men separated for neurosis or psychosis indicating that they

have had difficulty obtaining employment because of the nature of their discharges. Furthermore, the VA records and questionnaires contain considerable evidence that the VA frequently gave special employment assistance to these veterans. As a result, a number appear to have been hired for positions in industry as well as in government which they, in all probability, would have been unable to obtain if they had not been discharged with a service-connected disability. Certainly, the very marked differences in the postwar adjustment patterns of the neurotics and psychotics, both of whom received the same type of discharge, are not readily explainable in terms of restricted occupational opportunity due to conditions of separation from military service.

Although the high incidence of unemployment among the psychotics results in part from the fact that a number of these men were undergoing long-term hospitalization at the time of follow-up, this is by no means the only determinant. Slightly over 65 per cent of the psychotic unemployed were not in hospitals. Among the neurotics 83 per cent of the unemployed were not institutionalized at follow-up. None of the controls were in hospitals. It is also clear from the analysis of occupational level at the time of last known employment that the majority of the unemployed psychotics last worked at Level 4, unskilled jobs. This appears to be true of 75 per cent of the psychotics who were not working at follow-up. The figure for the neurotics is 61 per cent, but of course much fewer were out of work. Because of the low number of unemployed controls and the difficulties involved in applying weighting procedures to such a small sample, a reliable figure cannot be given for this group, but it is almost certainly not over 50 per cent, and probably considerably below that.

This suggests that there are, among those with a history of psychosis, a large number of men who drift in and out of unskilled jobs, but spend most of their time in an unemployed status. If this is true, one would expect to find a group of individuals who "just happened" to be employed at the time of follow-up included in the Level 4 frequency. A check on this assumption reveals that almost 40 per cent of the psychotics who were employed in unskilled jobs at follow-up had a history of long periods of unemployment and frequent job changes. Among this latter group as well as among the unemployed psychotics who had last worked in Level 4 positions, the usual pattern was not a gradual or even a precipitous decline in employment level. Characteristically, these men had never worked in any job above the semiskilled level; perhaps the majority, never above the unskilled.

Within the limits of generalization possible from the present samples, several guidelines for personnel policy emerge. Although these conclusions cannot be assumed to have validity in all employment situations, they should in most instances contribute to the efficiency of a company's personnel procedures. Ideally, however, local studies, which take into account variations in the labor market as well as company hiring and promotion procedures, should be carried out before policy is established.

Men with a history of psychosis apparently divide into two groups of approximately equal size. The first, as far as is presently known, make quite satisfactory employees. Further study may indicate that in spite of the fact that they adjust at approximately the same occupational level as other men from similar backgrounds, they do so at the expense of somewhat poorer performance. There is at present, however, no evidence to support this conclusion. The second group, on the other hand, appears to represent a poor employment risk. These men, whose histories may be expected to contain evidence of considerable unemployment, are not likely to stay on the job very long—either they will have to be released or they will leave on their own volition. They are very likely to have held only unskilled positions previously. Any man with a history of psychosis, who has worked at unskilled jobs in the past and who has been unemployed over considerable periods of time, should, under economic conditions similar to those of the postwar period, be hired only with the full realization that there is a high

probability of his soon becoming unemployed again.

By way of contrast, the neurotics as a group would appear to represent a much better employment risk. They are not likely to spend long periods among the ranks of the unemployed. However, men with a history of neurosis cannot be expected to attain supervisory and skilled positions as frequently as other employees. An employer who hires such a man with the expectation of moving him into a higher level position is likely to be disappointed. This does not imply that the former neurotic is necessarily a poor worker. The limited evidence we have suggests that as a group neurotics are likely to perform just as well as other employees, at least in lower level jobs (1). It seems probable, however, that they do typically lack whatever qualities make for placement or continued employment in supervisory and skilled positions.

Summary

Employment status approximately 10 years after discharge was determined for representative samples of men inducted into the army during late 1942 and prematurely separated for neurosis and psychosis. Similar information was obtained on a control group matched in terms of education, age, race, and county of residence. It was found that nearly half of the men with a history of psychosis were unemployed at follow-up as compared with approximately seven per cent of the controls. The neurotics did not have a reliably higher incidence of unemployment, but did tend to concentrate in low level occupations, particularly those of an unskilled nature. The proportion of men working in skilled and supervisory positions among the neurotics was considerably below that of the controls. This downward shift in occupational level was not present among those psychotics who were employed.

Received January 2, 1958.

References

1. Markowe, M. Occupational psychiatry: An historical survey and some recent researches. *J. ment. Sci.*, 1953, **99**, 92–101.
2. Miner, J. B. *Intelligence in the United States.* New York: Springer, 1957.
3. Miner, J. B., & Anderson, J. K. Intelligence and emotional disturbance: Evidence from Army and Veterans Administration records. *J. abnorm. soc. Psychol.*, 1958, **56**, 75–81.
4. Tomkins, S. S., & Miner, J. B. *The Tomkins-Horn Picture Arrangement Test.* New York: Springer, 1957.

Journal of Applied Psychology
Vol. 42, No. 5, 1958

Adequacy of the Residual Sensory Cues for Psychomotor Performance of Arm Amputees [1, 2]

Hilde Groth and John Lyman

University of California, Los Angeles

Amputation of the upper extremity deprives the individual not only of much of his normal motor function but also of a large amount of the sensory information which regulated this motor function. Although artificial arms constitute a fairly efficient mechanical replacement for some of the incurred motor loss, experimental results indicate that performance of visuo-motor tasks with a prosthesis shows a considerable decrease of performance speed (3). For the purpose of future design modifications it is of practical interest to ascertain whether this performance decrement should be attributed primarily to the loss of sensory information and, hence, to inadequate sensory regulation of the prosthesis movements or to a relatively inadequate motor replacement by the prosthetic device. Both of these possibilities focus on a third alternative, an increase in "central integration time." For the amputee any psychomotor task is necessarily much more complex than for the nonamputee since the prosthesis is interposed between the muscular motion and the control operation as an additional technical link. Aside from mechanical delays in the device itself, coordination of this extra link into the operating cycle conceivably can prolong the reaction time.

The specific purpose of this study was to investigate the extent of the usefulness of remaining sensory cues for prosthesis performance on two visuo-motor tasks. The sensory cues involved consist of a composite of afferent information received from the pressure of the socket on the stump, from the pressure exerted by harness straps, from movements of the stump and shoulder girdle during prosthesis operation and, in certain cases, from movements of and pressures within the biceps cineplastic tunnel as well as an undefined amount of visual information.[3] For laboratory investigation of the role of these cues a manipulation and tracking task was chosen to provide information about the following variables:

Task 1—Manipulation Task

a. Adequacy of object weight perception with pressure cues.
b. Adequacy of response speed on a task requiring gross motions with the prosthesis.
c. Problem of "central integration time" in a complex task.

Task 2—Tracking Task

a. Adequacy of pressure cues for dynamic force adjustments.
b. Adequacy of response speed when no gross motions of stump and prosthesis are required.
c. Problem of "central integration time" in a complex task.

Accuracy of weight perception during manipulation was considered to be a fair criterion for the adequacy of the available sensory cues. An excessively high level of grasp force during manipulation in comparison to the level of force exerted by nonamputees on the same task would be indicative of too little "feel" in the prosthesis. Accuracy of dynamic pressure discrimination, i.e., performance of a task requiring continuous adjustments of grasp force without change in terminal device opening, by exclusion of gross motions, was ex-

[1] This work was supported by Contract No. VAm-23110 between the Veterans Administration and the Department of Engineering, University of California, Los Angeles. The opinions expressed are those of the authors and do not necessarily represent those of the Veterans Administration.

[2] The authors wish to thank C. L. Taylor, Project Leader, for his cooperation in permitting this research to be undertaken, and J. R. Zweizig for the development of the electronically controlled apparatus.

[3] A cineplastic tunnel is a skin-lined tunnel constructed surgically in the biceps or other muscle and used as a "muscle motor" to power the control cable for prehension. A description has been published elsewhere in this journal (3).

pected to provide an estimate of the pressure sensitivity of the cineplastic tunnel or alternatively of the shoulder girdle areas underneath the harness straps.

Procedure

Subjects

Group I: 10 right-handed engineering students performing with the right hand.

Group II: 10 right-handed engineering students performing with the left hand.

Group III: 8 unilateral below elbow amputees wearing harness-controlled prostheses, performing with the hook.

Group IV: 4 unilateral below elbow amputees wearing cineplasty-controlled prostheses, performing with the hook.

Group V: 3 unilateral above elbow amputees wearing harness-controlled prostheses, performing with the hook.

All amputees were regular prosthetic hook wearers and had served repeatedly as research amputees in the laboratory. The students were recruited from classrooms.

Performance of unilateral amputees with the artificial limb was considered as performance with the nonpreferred limb regardless of which side had been amputated.

Apparatus and Tests

Manipulation apparatus. The electronically controlled instrument has been described elsewhere (**2**). Its main components consisted of a 3×6 control board matrix, a 3×6 lightbank display matrix, and a thin-walled split aluminum cylinder force transducer instrumented with Baldwin Lima SR-4 strain gages. The cylinder had a diameter of 5.0 cm., a slot opening of 1 mm., and weighed 422 gms. It was designed to assure nearly isometric muscle contraction during prehension. The apparatus permitted simultaneous recording of the integral of prehension force over the duration of the trial and of the total manipulation time for the trial length. Mean prehension force and time per transport were taken as criterion measures. Task performance required only the simple motion elements of grasp, transport and release. Task complexity was changed in terms of pace of performance and pattern of motion. The arrangement was placed on a table 30" high.

Tracking apparatus. Details of the instrument have been described in detail elsewhere (**2**). A "random" target problem was generated on a dual beam oscilloscope in the form of a vertically moving line across the face. The line was tracked with a light dot controlled by prehension force variations. The same split cylinder force transducer as was manipulated in Task 1 was placed on a support and used as a controller. The target line had a maximal vertical motion range of ± 3.0 cm. from the center of the scope face. The tracking dot had its zero position 3.0 cm. below the center and had a maximal vertical range of +6.0 cm. The force required for tracking with this dot was approximately 165 gm./cm. The apparatus was placed on a table 30" high; the force transducer rested on a 10" stand in front of the scope. Number of target hits, time on target and time on target per hit contributed the criterion measures.

Terminal devices for amputees. A left or right APRL voluntary closing hook, modified by removal of the locking mechanism, was worn by each amputee. These hooks have been described previously (**3, 4**). Voluntary closing hooks were chosen because the amputee had to apply control cable tension in direct proportion to the prehended load. The ratio between output force at the hook tip and control cable force was measured as 2.5 : 1. Prehension force of 1,000 gm. thus required a cable pull of about 2,500 gm.

The hook tips were covered with commercial rubber sleeves in order to obtain similar conditions of surface friction for bare hand and hook performance. The coefficient of static friction between skin and aluminum was approximately 0.7 and for the rubber sleeves it was 0.5.

Experimental noise. Since there was a possibility that amputees would respond to some auditory cues from the prosthesis, head phones were worn by all *Ss*. A random noise generator provided the masking noise. The head phone output was adjusted to 90 db SPL.

Routine

The following six treatments were administered in a subject by treatment design during a single experimental session:

1. Pursuit tracking task; duration: 5 min.
2. Manipulation task, self-paced, regular pattern; duration: 10 min.
3. Manipulation task, self-paced, random pattern; duration: 10 min.
4. Manipulation task, experimenter-paced, regular pattern; duration: 10 min.
5. Manipulation task, experimenter-paced, random pattern; duration: 10 min.
6. Manipulation task, speed trial.

Treatments 2, 3, 4 and 5 were given in random order.

Tracking task. The cylinder was resting on a support in front of the oscilloscope and *S* pursued the moving target line with the light dot for five minutes.

Manipulation task. 1. Pace of performance: (*a*) self-paced (SP): *S* was asked to adjust his pace so that he could perform the task comfortably as an eight-hour job. (*b*) Experimenter paced (EP); the stimulus lights changed at 2-sec. intervals.

2. Movement pattern. (*a*) Regular pattern. The *Ss* started at the lower left side of the control board and placed the cylinder into the adjacent hole, working back and forth, toward and away from the body. Under the EP condition, the signal for each move

Table 1

Mean Prehension Forces and Variabilities on Manipulation Task

Treatments	Nonamputees		Amputees	
	$\bar{X}$ PF (gm.)	PF s (gm.)	$\bar{X}$ PF (gm.)	PF s (gm.)
Self-paced Regular pattern	1,022	561	1,478	769
Self-paced Random pattern	1,174	617	1,359	558
Experimenter-paced Regular pattern	1,213	254	1,328	522
Self-paced Speed trial	1,305	898	1,332	519

Note.—Just slip force: 1. bare hands ~604 gm., 2. hooks ~844 gm.

was given by the flashing of a small light bulb fastened in the center of the upper border of the display panel. (*b*) Random pattern. The *Ss* placed the cylinder into that hole of the control panel which corresponded in location to the lighted circle of the display matrix. In the SP condition, the display light would change to the next position as soon as the contact of the preceding move was made. For the EP condition, the location of the stimulus light would change at two seconds' intervals.

3. Speed trial. The *Ss* were asked to perform five rounds of the regular pattern as fast as possible. No head phones were worn for this condition.

Each task was explained in detail and demonstrated by *E* before the experimental run.

Nonamputee *Ss* grasped the cylinder around the upper rim with the pads of all five fingers leaving the slot between the thumb and index finger. Amputee *Ss* grasped it with the length of the rubber covered hook tips with the slot pointing away. By this method we were able to approximate the sizes of the contact areas of bare hand and hook performance to satisfy the equation, $F = P \times A$, where F = force, P = pressure, A = area of contact.

All treatments were administered while the *Ss* were standing.

Calculations. All comparisons for statistical inference were made by nonparametric methods since data of amputee performance showed neither normality of trait distribution nor homogeneity of variance (**8**). The significance level was set at $P \leq .01$.

Table 2

Mean Times per Transport and Variabilities in Manipulation Task

Treatments	Nonamputees		Amputees	
	$\bar{X}$ T/tr (sec.)	T/tr s (sec.)	$\bar{X}$ T/tr (sec.)	T/tr s (sec.)
Self-paced Regular pattern	.55	.14	1.01	.32
Self-paced Random pattern	.70	.21	1.20	.36
Experimenter-paced Regular pattern	.60	.12	.90	.16
Self-paced Speed trial	.40	.09	.78	.23

Table 3

Mean Tracking Performance and Variability of Amputees and Nonamputees

Criterion Measure	Amputees		Nonamputees	
	$\bar{X}$	s	$\bar{X}$	s
Total number of hits	241.14	68.92	302.75	39.49
Number of hits—first minute	49.71	16.00	62.30	10.91
Number of hits—fifth minute	48.36	14.63	62.15	8.38
Time on target—first minute	21.93	8.07	31.50	3.24
Time on target—fifth minute	22.29	11.13	28.20	5.51
Time on target per hit—first minute	0.48	0.26	0.53	0.13
Time on target per hit—fifth minute	0.57	0.55	0.47	0.14

Results [4]

Manipulation Task Comparison of Prehension Forces (PF)

Comparisons of amputee with nonamputee performance on this task were made between all amputees (Groups III, IV, V) and the 10 nonamputee *S*s (Group II) performing with the nonpreferred hand.

Inspection of the mean PF values for four treatment conditions for these groups showed consistently higher force levels for the amputees. However, statistical comparisons of the median differences by the Mann-Whitney *U* test failed to reach significance.

No data could be reported for the experimenter-paced random pattern condition since none of the amputees were able to complete this task. Complexity of the task led to confusion and performance breakdown in each case.

Mean values and variabilities are shown in Table 1.

Manipulation Task: Comparison of Times Per Transport (T/tr)

Statistical comparison between amputee and nonamputee performance by the Mann-Whitney *U* test showed that amputees performed consistently slower. This difference was significant for each of the four conditions. The mean values and variabilities are reported in Table 2.

Manipulation Task: Performance of Amputee Subgroups

The small size of the amputee subsamples did not permit any statistical comparisons of performance. However, inspection of the raw data indicated a lack of performance differences due to amputation level or method of muscle-prosthesis coupling.

Tracking Task: Comparison of Target Hits

Tolerance limits for "on target" were set as a 6 mm. wide band on the scope face. Visually this meant for the *S* that he was on target whenever the perimeter of the tracking dot touched the target line.

Comparisons of amputee with nonamputee performance have been made between all amputees (Groups III, IV, V) and all nonamputees (Groups I, II). Combining Groups I and II seemed justified since none of the *S*s had any previous experience with a similar task.

Statistical comparisons were made for three time intervals: (*a*) total number of hits during the five-minute trial; (*b*) number of hits during the first minute; (*c*) number of hits during the fifth minute. None of these comparisons reached statistical significance.

Tracking Task: Comparison of Time on Target

Comparisons have been made for the first and fifth minutes of performance. None of the differences reached statistical significance.

[4] The statistical tables have been deposited with the American Documentation Institute. Order Document No. 5690 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Tracking Task: Comparison of Time on Target Per Hit

Comparisons for first and fifth minute of performance again failed to differentiate between amputees and nonamputees.

Table 3 summarizes the mean values and variabilities for the tracking task.

Tracking Task—Performance of Amputee Subgroups

No statistical comparisons were possible because of the small size of the subsamples. Inspection of the raw data again failed to indicate pronounced performance differences attributable to amputation level or method of coupling. However, one of the *Ss* with cineplastic prosthesis control could not complete the task because of a tremor in the biceps.

Discussion

To the authors' knowledge, no systematic investigation of the role of cutaneous and kinesthetic information during active motor performance with artificial arms has been conducted previously. The present study also falls far short of the goal of assessing the importance of the various available cues in psychomotor skills. For example, it would have been desirable to study performance changes during selective anesthetization of tactile and proprioceptive afferents without also affecting the motor efferents. The authors made some preliminary attempts to do this and failed, for present methods of anesthesia do not seem to permit adequate separation of sensory components.

The problem, therefore, has been approached on a less refined level of analysis and the results provide at best an initial evaluation of the practical utility of the composite of all residual cues.

The principal findings of this experiment can be summarized as follows:

1. Amputees, regardless of amputation level and type of prosthesis control, made adequate use of the available cues. This was evidenced by prehension force adjustments when manipulating an object of unknown weight as well as by the tracking scores which also were dependent on the accuracy of pressure discriminations.

2. Amputees showed a pronounced impairment in response speed when the task involved gross motions as was shown on the manipulation task. Response speed on the tracking task, which did not involve gross movements with the prosthesis, seemed to be comparable to normal hand performance.

3. Task complexity requiring additional "central integration time" as accomplished by imposition of a random pattern upon the experimenter paced manipulation task led to breakdown of performance for all amputees.

From these results we inferred that amputees apparently have fairly adequate sensory control for execution of simple visuo-motor tasks. However, visuo-motor tasks requiring gross motions show impairment of amputee performance due to an increase in time needed for completion of a given movement with the prosthesis. This was brought out when speed stress was increased by changing the task complexity.

The present investigation does not permit us to estimate the relative importance of vision to the reliable interpretation of the cutaneous and proprioceptive cues perceived during performance. It seems to be fairly well established that vision alone in the absence of other sensory information cannot serve as an adequate source of information for motor performance (**6**, **7**). This would be expected to be particularly the case where the task requires accurate prehension force control as was true in our tracking task. On the other hand it has been found that blind-folded performance will cause a decrement in manipulation as well as in travel time, and even after prolonged practice these components of performance will not reach the level of proficiency attainable with unrestricted vision (**5**). An experiment in this laboratory which compared a prosthesis of the type used here with a prosthesis in which prehension force was electrically actuated by an on-off switch showed an increase in reaction time only for the electric arm when visual cues were removed (**1**). Because of the type of actuation fewer nonvisual cues were available for the electric arm.

The present results as well as some of our earlier ones (**1**, **4**) seem to indicate clearly that amputees are capable of using nonvisual

as well as visual cues effectively with present prosthesis designs. Accordingly, we feel that no specific sensory aids should be recommended for prosthesis design until both the adequacy and reliability of nonvisual cues have been studied for tasks in which the obstruction of vision has been investigated further.

Summary

This investigation was designed to evaluate three possible sources of performance decrements of amputees: (*a*) Inadequate sensory information during prosthesis performance. (*b*) Increase in duration of the mechanical motions in comparison to normal movements. (*c*) Increase of the "central integration time" during a task due to the complexity added by the prosthesis. Identification of the principal source of decrement would serve as an indicant for the site of modification of the engineering design of the prosthesis. This modification could take the form of sensory aiding devices, improvement of the mechanical function, or simplification of the coupling to the control musculature.

Two visual motor tasks were chosen to provide information about these possible sources of performance deterioration: (*a*) Manipulation task; (*b*) pursuit tracking task. Twenty nonamputees and 15 unilateral amputees served as experimental *S*s. The amputees consisted of eight below-elbow amputees with harness control, four below-elbow amputees with cineplastic control, and three above-elbow amputees with harness control.

The principal findings were: (*a*) Performance showed that all amputees received adequate pressure cues; (*b*) all amputees showed a pronounced impairment in response speed when the task required gross movements with the prosthesis; (*c*) task complexity inducing time stress by requiring additional "central integration time" led to performance breakdown and confusion.

Since the study did not permit an estimation of the adequacy and reliability of cutaneous and proprioceptive cues in the absence of vision, another investigation in which vision is obstructed has been suggested.

Received January 2, 1958.

References

1. Gottlieb, M., Santschi, W., & Lyman, J. Evaluation of the Model IV-E2 electric arm. Department of Engineering, Univer. California, Los Angeles, 1953. (Rep. 53-23.)
2. Groth, Hilde. An experimental assessment of prehension force as a measure of effort in psychomotor skills. Unpublished doctoral dissertation, Univer. California, Los Angeles, 1957.
3. Groth, Hilde, & Lyman, J. Relation of the mode of prosthesis control to psychomotor performance of arm amputees. *J. appl. Psychol.*, 1957, **41**, 73–78.
4. Groth, Hilde, & Lyman, J. A comparison of two modes of prosthetic prehension force control by arm amputees. *J. appl. Psychol.*, 1957, **41**, 325–328.
5. Huiskamp, J., Smader, R. C., & Smith, K. N. Dimensional analysis of motion: IX. Comparison of visual and nonvisual control of component movements. *J. appl. Psychol.*, 1956, **40**, 181–186.
6. Lassek, A. M. Effect of combined afferent lesions on motor function. *Neurology*, 1955, **5**, 269–272.
7. Twitchell, Th. E. Sensory factors in purposive movement. *J. Neurophysiol.*, 1954, **17**, 239–252.
8. Walker, H. M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

Journal of Applied Psychology
Vol. 42, No. 5, 1958

Curriculum Assessment with Critical Incidents

Albert S. Glickman

U. S. Naval Personnel Research Field Activity, Washington [1]

and T. R. Vallance

Human Resources Research Office

By and large, even in industry, adequacy of training programs has been assessed by subjective judgment rather than by research methods (5). As Flanagan (1, 2) has remarked, there is need to supplement the usual collection in conference of "leaders" or "experts" by systematic collection and analysis of factual data.

This investigation explores the use of "critical incidents" (1, 2, 3) for assessing the relevance of various aspects of a training program to the job performance requirements of its graduates. Specifically, it involves the relevance of subject matter taught in the Navy's Officer Candidate School (OCS) to the duty requirements of new officers aboard destroyer-type ships.

The 17-week OCS training aims to prepare the new ensign to profit from experience and rapidly achieve a satisfactory level of job performance. It is not intended to produce specialists. Rather, the purpose is to provide junior officers with the familiarity in subject matter areas containing technical problems needed to develop an understanding of a variety of technical functions, and to develop ability to supervise technicians.

The critical incident procedure, as adapted here, takes cognizance of the supervisory nature of the junior officer's duties and at the same time permits the use of duty performance behaviors of sufficient specificity to indicate OCS curriculum elements most and least relevant to significant shipboard duties. Also, by concentrating upon observed actual performances on duty, the procedure makes it possible to identify duty areas which are not covered by instruction, thus contributing to a further definition of what the objectives and content of training should be.

The essential questions to which this research is addressed are these: "What are the things which happen frequently to ensigns, or which their superiors think they ought to be able to handle soon after reporting aboard?" and then, "Are these the things to which attention and emphasis are also given by OCS instruction?"

Our legitimate concern is exclusively with the performance requirements (and relevant training) reflected in the content of incidents, whether they be found in reports of "effective" or "ineffective" incidents. This research does not tell us how much OCS training contributes to causing effective or ineffective incidents.

Determining the Gross Relevance of OCS Curriculum to Duty Incidents

Basic Data

An earlier study (6) had accumulated incidents involving destroyer officers from all training sources. Of those, 1,073 incidents reported on general line ensigns constituted our pool of basic data. This sample of incidents is used here to provide a description and operational definition of the critically significant elements of the duties of new officers aboard destroyers. Incidents were typed on IBM cards to facilitate handling and analysis.

Analysis of the Incident Pool with Reference to Curriculum Areas

The incident cards were submitted to subject matter specialists, OCS officer-instructors, who carried out a review sequence that ultimately required that each incident be identi-

[1] This research was conducted while the authors were with the Officer Personnel Research Program of the American Institute for Research, working under contract Nonr 890(01) with the Office of Naval Research. A more extensive report was prepared at that time (4). The viewpoints expressed herein are not to be construed as those of the U. S. Navy or U. S. Army.

fied with a lesson in the lesson-plan of one of the seven sections of the curriculum: 1. Seamanship, 2. Orientation and Military Justice, 3. Navigation, 4. Operations, 5. Naval Weapons and Fire Control, 6. Engineering and Damage Control, and 7. Military. Category 8, labelled "Other," was established for those incidents which reflected motivation and interest in the job rather than the possession of specific skills or knowledges.

The instructional set for sorting effective incidents concerned the curricular locus of skills, knowledges and attitudes shown in each incident. For the ineffective incidents the emphasis was on the skills, knowledges and attitudes which, if learned, could have prevented the ensign from getting involved in the incident.

Within each subject matter category, incidents were classified in two ways: "Taught" —the relevant skills, knowledges and attitudes are currently the subject of specific lessons of instruction; and "Not Taught"—pertains to material not covered, but which logically belongs in one of the school sections and probably would be taught were time and facilities available. All "Other" incidents were, of course, of the "not taught" variety.

No assumption can be made that individual incidents, or even incidents of a given sort, are equally significant from the operational standpoint. Yet, if a large number of the critical incidents reported fall into a given subject matter area, it is difficult to escape the conclusion that there is much important activity aboard ship to which that area of instruction is relevant. Likewise, if the number of incidents in certain instructional areas is small, unless it is found that the incidents are individually of vital significance, it would seem that the learnings in such areas are not heavily drawn upon in the performance of a new officer's duties.

It appears from Table 1 that it is from his background in the courses Orientation and Military Justice, Seamanship, and Operations, that on reporting to a destroyer the new ensign might expect to draw most frequently, while being less likely to be called upon to make use of what he learned in Navigation, Engineering and Damage Control, and Naval Weapons and Fire Control.[2]

Can the fact that 62.9% of the classifiable incidents are said by instructor-judges to be taught at OCS be considered to reflect a "good" state of affairs? Since the limits imposed by time and facilities have not been considered and since there is no standard of what "percentage taught" can be considered good, this question cannot be answered in an absolute sense. However, we might say that the higher the "percentage taught" the better.

Thus, it might be presumed that of the six academic courses, Seamanship and Orientation and Military Justice most completely cover their respective areas, and that the Naval Weapons course least adequately treats of the kinds of situations that destroyer officers report as critical.[3] However, there are instances where 50 or more incidents are considered relevant to one hour of instruction, leading one to discount the direct interpretation of "per centage taught" as a sufficient criterion of the adequacy of subject matter coverage.

References to the number of incidents assigned to specific hours of instruction, along with further examination of the contents of sessions by subject matter experts, would seem to provide a more meaningful employment of data such as these for purposes of curriculum analysis.

Developing an Index of Incident-Behavior "Importance for Early Usefulness"

Rationale

The main objective of the next phase was to obtain expert judgments regarding the sig-

[2] Some incidents were considered relevant to more than one lesson. However, in Table 1, no incident was counted more than once, since when an incident was identified with more than one lesson, these lessons were always in the same curriculum section. The judges could not classify 17 of the original 1073 incidents.

[3] If the "percentage taught" figures for effective and ineffective incidents for the six academic courses (all except Military) are arranged in rank order, the correlation between ranks for "effectives" and "ineffectives" is .79, suggesting that, whether relevant effective or ineffective incidents are used, the relative coverage by courses is much the same. Table 1 also reveals that overall the "percentage taught" is not much different for effective and ineffective incidents (64.6% vs. 62.1%).

Table 1

Results of Sorting Critical Incidents by Subject Matter Areas

	Effective		Ineffective		Summary	
	N	%[a]	*N*	%[a]	*N*	%[a]
1. Seamanship						
Taught	52	4.8	106	9.9	158	14.7
Not-Taught	12	1.1	40	3.7	52	4.8
Total	64	6.0	146	13.6	210	19.6
% Taught[b]	81.3		72.6		75.2	
2. Orientation and Military Justice						
Taught	125	11.6	240	22.4	365	34.0
Not-Taught	46	4.3	88	8.2	134	12.5
Total	171	15.9	328	30.6	499	46.5
% Taught[b]	73.1		73.2		73.1	
3. Navigation						
Taught	5	0.5	4	0.4	9	0.8
Not-Taught	0	0.0	1	0.1	1	0.1
Total	5	0.5	5	0.5	10	0.9
% Taught[b]	100.0		80.0		90.0	
4. Operations						
Taught	19	1.8	66	6.2	85	7.9
Not-Taught	22	2.1	57	5.3	79	7.4
Total	41	3.8	123	11.5	164	15.3
% Taught[b]	46.3		53.7		51.8	
5. Weapons						
Taught	7	0.7	8	0.7	15	1.4
Not-Taught	11	1.0	28	2.6	39	3.6
Total	18	1.7	36	3.4	54	5.0
% Taught[b]	38.9		22.2		27.8	
6. Engineering						
Taught	16	1.5	14	1.3	30	2.8
Not-Taught	4	0.4	18	1.7	22	2.1
Total	20	1.9	32	3.0	52	4.8
% Taught[b]	80.0		43.8		57.7	
7. Military						
Taught	0	0.0	2	0.2	2	0.2
Not-Taught	2	0.2	0	0.0	2	0.2
Total	2	0.2	2	0.2	4	0.4
% Taught[b]	0.0		100.0		50.0	
8. Other						
Taught	0	0.0	0	0.0	0	0.0
Not-Taught	26	2.4	37	3.4	63	5.9
Total	26	2.4	37	3.4	63	5.9
% Taught[b]	0.0		0.0		0.0	
Total						
Taught	224	20.9	440	41.0	664	61.9
Not-Taught	123	11.5	269	25.1	392	36.5
Total	347	32.3	709	66.1	1056	98.4[c]
% Taught	64.6		62.1		62.9	

[a] Percentages in these columns are based on the original total number; *N*/1073 rounded to nearest tenth.

[b] Percentages in these rows represent the number of "taught" incidents in relation to the total number of incidents within respective subject categories (i.e., "taught"/"total").

[c] Seventeen incidents (1.6%) were incapable of being judged with confidence and are not included in this summary.

nificant aspects of duty for which training should prepare the new ensign.

To place the notion of importance clearly in the framework of training, it was reasoned that to the extent that preduty training provides the trainee with skills, knowledge, or attitudes required of him in the first few months of duty, or a sound background on which experience will develop them, that training is successful. To the extent that it concentrates on that which the trainee will not be called upon to use until after he has had a year or more of duty, or has been sent to other schools to acquire or reacquire them, initial training is being misdirected. From this standpoint, the immediacy of the opportunity or requirement to handle any given situation would be an index of its "importance." It is in this sense that "importance," standing for "importance for early usefulness," is used in this article.

Procedure

A questionnaire, the Junior Officer Training Requirements Checklist (JOTRC), was prepared in 10 forms containing an aggregate of 985 items.[4] The final format evolved from four preliminary try-out versions.

Each form listed approximately 100 incidents, and was divided into two parts. Part I contained effective incidents and Part II ineffective incidents for Forms 1, 3, 5, 7, and 9. For Forms 2, 4, 6, 8, and 10, the order was reversed. About one hour was needed to fill out a form. The incidents related to all OCS instructional areas, and included those judged earlier to be among the "not taught" as well as those included in the curriculum.

The specific questions were:

1. For the situations involving effective incidents: "How soon after reporting aboard would you expect the new officer to be able to do this?"

2. For the situations involving ineffective incidents: "How soon after reporting aboard would you expect the new officer to be able to handle this situation satisfactorily?"

The directions specified the writing of the number of months (directly after being commissioned), ranging from 0 (for judgments that the reserve officer should be able to perform well immediately on reporting aboard) to 12. "X" was to be used if the respondent would not expect the new officer to be able to handle the given type of situation satisfactorily until after he had been aboard for more than a year. "N" was to be used if the reporter felt unable to make a judgment because of wording or interpretation of the incident items.[5]

Two copies of the checklist were sent to each of 170 destroyer-type ships, one for the commanding officer (CO) and one for his executive officer (XO). Each of the 10 forms was sent to 30 to 50 officers. A total of 301 JOTRCs were returned in time to be analyzed.

For each of the 985 incidents, the estimates obtained from COs and XOs were tallied and a median value was found. We call these medians TESP ("time expectancy for satisfactory performance") values.

Analyses of Results

Reliability of TESP Values

Using Form 8 of the JOTRC, a reliability estimate for the CO–XO TESPs was obtained. The 44 questionnaires returned (out of 50) were divided in two groups of 22 each. For each item two TESPs were obtained.[6] Correlation of the two sets of TESPs over the 95 items gave a coefficient of .93.

Distribution of TESP Values

The distribution of TESPs for the 985 JOTRC items was divided into incidents of "effective" (E) and "ineffective" (I) performance.

There was a distinct difference in TESPs assigned to E and I performances (median values of 6.0 and 2.5 months, respectively). The officers responsible for operating these ships were most immediately concerned with keeping the novice (and themselves) out of trouble, and were content to wait longer for a quality of performance exceeding what is normally considered satisfactory. Overall the TESPs tended to pile up at the bottom (more "important") end of the scale (Table 2).

These estimates seem to bear out assertions that destroyer duty requires of an officer rapid learning and assumption of responsi-

[4] This number is smaller than that involved in the two instructor sorts because of deletion of incidents: (*a*) of clear duplication; (*b*) which might be misinterpreted by the judges because the situations represented contraventions of Navy policy (e.g., a number of incidents originally submitted as "effectives," turned out to be violations of regulations).

[5] The results to be presented must be interpreted in terms of the *status quo* at the time of this investigation. To the extent that the operational, training, or personnel factors in the Navy change conditions, the values obtained might be altered.

[6] In all computational work the "X" response, meaning "more than 12 months," was assigned a value of 13; the lower the value, the greater the relevance of the incident to the OCS curriculum.

Table 2

Frequency Distribution of Incidents by Curriculum Area, Subcategorized as "Taught" and "Not Taught," With Mean of TESP Values for the Incidents in Each Subcategory Indicated

Curriculum Area	Taught			Not Taught			Total		
	No. of Incidents	% of Incidents[a]	Mean TESP	No. of Incidents	% of Incidents[a]	Mean TESP	No. of Incidents	% of Incidents[a]	Mean TESP
1. Seamanship	143	16.3	4.34	46	5.2	4.28	189	21.5	4.32
2. Orientation and Military Justice	300	34.2	2.73	99	11.3	2.92	399	45.4	2.78
3. Navigation	8	0.9	5.13	1	0.1	2.00	9	1.0	4.78
4. Operations	77	8.8	4.48	67	7.6	5.00	144	16.4	4.72
5. Naval Weapons	12	1.4	7.30	38	4.3	6.26	50	5.7	6.51
6. Engineering and Damage Control	26	3.0	5.46	18	2.1	5.50	44	5.0	5.48
7. Military	2	0.2	1.00	2	0.2	7.00	4	0.5	4.00
8. Other	—	—	—	39	4.4	4.32	39	4.4	4.32
Summary	568	64.7	3.62	310	35.3	4.33	878	100.0	3.87

[a] Percentages based on 878 incidents; *N*/878. Eliminated, because more than 10% of respondents omitted an answer or indicated that they could not clearly interpret the incident, were 91 of the original 985 items. Not included because they could not be judged with confidence as to area of relevance were 16 incidents.

bility involving manifold skills and knowledges, even allowing for the fact that all ensigns are not expected to become competent in all of the 985 incident situations.[7] The fact that the COs and XOs expected the ensigns to be able to handle a sizable majority of the situations after only a few months on board seems to lend corroboration to the notion that the over-all adequacy of the curriculum is, in general, proportionate to the coverage given to these situations, even allowing for differences in TESPs of different incidents. In other words, if the relevant skills, knowledges, and attitudes are "taught," the curriculum conforms to CO–XO judgments. If these are not taught, COs and XOs say, in effect, they ought to be.

TESP Values of Incidents Assigned to Subject Matter Areas

The next analyses deal with the relative "importance" of incidents assigned to different subject matter areas and the relation between the number of incidents found to be relevant to a given area and the TESP values assigned to those incidents.

In order to provide a basis of interpretation least vitiated by ambiguity, the following analyses were restricted to 878 of the original 985 incidents included in the 10 JOTRC forms. Those 91 items, which more than 10% of respondents did not answer or indicated that they could not clearly interpret, were deleted. Sixteen items, which the OCS experts had not been able to assign with confidence to relevant subject matter areas, were also removed. The numbers of incidents "taught" and "not taught" in the six academic courses and in military training were counted, and the mean TESP was calculated for each of these categories and subcategories, as well as the "Others," as shown in Table 2. Among the academic courses there is a close correspondence between the number of incidents said to be relevant to a given course by OCS instructors and the mean TESPs derived independently from the judgments of "importance for early usefulness" by the COs and XOs. The three academic courses which embrace the largest number of incidents are also, in the same order, the courses which are the most "important" as indicated by the mean

[7] As indicated in the questionnaire instructions, the respondent is required to "Suppose [with respect to each incident] that his [the ensign's] responsibilities as an officer will give him the opportunity to handle the situation as shown."

TESP values. Of the six, only the courses in Navigation and Naval Weapons, the former with a small number of cases, are out of order with respect to these two quantities. Again it would appear that there is substantial parallelism between number and TESP value of incidents when sorted into OCS curriculum areas.

Some indication of the curricular validity of OCS academic offerings is shown by the fact that the mean TESP for material "taught" is lower (more important) than for material "not taught." However the caution introduced earlier, because of the large proportion of incidents said to be related to the material taught in a relatively small number of sessions, still applies.

We have gone one step further and pursued the analysis of the 568 incidents said to be "taught" in the sample of 878 incidents. We determined the mean TESP for the incidents assigned to each instructional session to which one or more incidents were judged to be relevant. A distribution of these 157 sessions against mean TESPs of incidents assigned to each session made it apparent that the lessons covering the largest number of relevant incidents also tend to be considered more "important for early usefulness" aboard ship. Using 28 lessons for which somewhat more reliable mean TESPs are available (based on 10 or more incidents) we again established a fair correspondence between "number" and "importance" of incidents in a given category. The correlation for these 28 lessons was .65.[8]

An inspection was made of the content of these 28 lessons. It supports the view that of all his background, his skills in human relations will be those which a new officer will be most immediately and most frequently called upon to demonstrate. It appears that the first thing a junior officer has to learn is how to manage his relations with contemporaries and seniors. Leadership and management of subordinate personnel show up as next most "important" skills. Demands upon his technical skills and knowledges are not made until later.

[8] The sign of the coefficient has been reversed here, to account for the fact that the more "important" items have lower TESPs, and thus permit direct reading and interpretation.

Discussion

It has been pointed out that the general purpose of OCS training is to develop officers with general familiarity in subject matter areas containing technical problems and with ability to supervise technicians, not technical specialists. It would appear that officers responsible for operating destroyers are in general agreement with that emphasis. They are perhaps inclined to give greater emphasis to the requirement that he be able to take up quickly, and carry on satisfactorily, responsibilities requiring "know-how" in dealing with people rather than technical skill. Critical incidents involving such skills are also, by far, the kind most frequently reported about ensigns.

Other uses of the results are available. For example, we have concentrated comment on the curriculum areas or lessons most frequently the subject matter of relevant incidents and to which most "importance" is attributed. We have not spoken much of the lessons involving relevance with none of our pool of incidents or those of low "importance." Obviously, these would be examined first in terms of their contribution to the training program and for the possibility of using the time given them for instruction on some of the incident-behaviors which were highly concentrated in a few sessions of the program studied (e.g., leadership).

The scope of this research has been limited. Although the feasibility of the method has been demonstrated by embracing the whole curriculum as the field of inquiry, insufficient data could be brought to bear on many areas. More concentrated attacks can be made upon specific curriculum problems, using the same procedures. For instance, if it were desired to revamp the Navigation offerings, critical incidents involving navigation could be requested from officers on ships and then sorted by lessons of relevance. Since number and "importance" of incidents tend to be associated, it might be possible to eliminate the determination of TESPs, if one were willing to sacrifice some precision.

Naturally, if any considerable changes in the curriculum were to be made, follow-up studies should be conducted to see if the de-

sired changes in on-the-job performance came about.

In this exploratory research we have not concerned ourselves with such matters as the policies, directives, costs and laws governing operation of a school, nor with the time, staff, techniques, or facilities, necessary or available to achieve satisfactory levels of competence through training in the various aspects of a curriculum. However, they are within the purview of the administrative authorities governing a training program.

The case in point here has been the Navy's Officer Candidate School. However, the applicability of the methodology to other training and educational settings is apparent.

Summary

This research sought to identify those aspects of the Navy OCS curriculum which were most and least relevant to duties of newly commissioned ensigns aboard destroyers; thus to provide information useful in improvement of training.

More than 1,000 critical incidents of effective and ineffective performance by destroyer ensigns were available from earlier research. These were sorted according to the OCS curriculum area to which each was most relevant. The "relevant" incidents assigned to each area, were then classified as: (*a*) "Taught"—currently the subject of specific lessons of OCS instruction, and (*b*) "Not Taught"—pertains to a subject matter area in the curriculum, but not covered due to time or facility limitations.

A Junior Officer Training Requirements Checklist was sent to 340 commanding and executive officers of destroyer-type vessels, and was completed by more than 300 of them. The checklist was prepared in 10 forms, each containing approximately 100 incidents. Each form was sent to 30 to 50 officers, with instructions to make a judgment for each incident as to: "How soon [in number of months] after his reporting aboard [directly after being commissioned], under normal conditions, would you expect the new [reserve] officer to be able to handle the situation to your satisfaction?" From their answers an index of "time expectancy for satisfactory performance" was determined for each incident with high reliability. It was assumed that the sooner the ensign is expected to handle a situation satisfactorily, the more "important" it is that the relevant material be learned at OCS.

The findings indicate that the new ensign most frequently and most immediately will be called upon to draw on background relevant to human relations, leadership, and personnel administration skills; technical skills are expected to be developed later.

Suggestions are made for utilization of the current findings and directions for further research.

Received January 9, 1958.

References

1. Flanagan, J. C. Research techniques for developing educational objectives. *Educ. Rec.*, 1947, **28**, 139–148.
2. Flanagan, J. C. The critical requirements approach to educational objectives. *Sch. & Soc.*, 1950, **71**, 321–324.
3. Flanagan, J. C. The critical incident technique. *Psych. Bull.*, 1954, **51**, 327–358.
4. Glickman, A. S., & Vallance, T. R. An exploratory study of the applicability of critical incident techniques to the assessment of curricula for officer candidate training. *USN Bur. Naval Pers. Res. Tech. Bull.*, 1954, No. 54-23.
5. Mahler, W. R., & Monroe, W. H. How industry determines the need for and effectiveness of training. *USA Personn. Res. Sect. Rept.*, 1952, No. 929.
6. Vallance, T. R., Glickman, A. S., & Vasilar, J. N. Critical incidents in junior officer duties aboard destroyer-type vessels. *USN Bur. Naval Pers. Res. Tech. Bull.*, 1954, No. 54-4.

Journal of Applied Psychology
Vol. 42, No. 5, 1958

A Personality Inventory Employing Occupational Titles [1]

John L. Holland

National Merit Scholarship Corporation

In the last two decades the research literature concerning the meaning of occupational choice—the relation of vocational interests to attitudes and personality—has implied with increasing clarity that a useful personality inventory might be constructed from occupational or interest test content (3, 4, 12, 15). The present experimental inventory represents a further testing and application of this hypothesis.

Essentially, the Holland Vocational Preference Inventory is a personality inventory which uses occupational titles for content. It is constructed to give a maximum amount of reliable and valid information with a minimum amount of testing and scoring time, skill, and expense. Many variables usually included in interest and personality tests have been integrated in the HVPI. These variables are assumed to yield a broad range of information concerning the *S*'s personal adjustment, values, attitudes, and vocational motivation. The inventory is self-administering. The *S*s record their "feelings and attitudes" about occupational titles by indicating either "interest" (appeal) or "dislike" for each of the 300 items.

Because of the relative neutrality of its title and occupational content, the inventory appears to have a number of uniquely desirable qualities. Occupational titles provide a subtle set of stimuli which minimize the negative reactions sometimes provoked by more obvious personality inventories, reduce the *S*'s need to "fake" since the inventory is perceived as a "vocational" inventory, and virtually eliminate those requests for personality interpretations which usually follow on the heels of the administration of other personality inventories.

[1] I wish to acknowledge the support and encouragement of numerous colleagues and students at Western Reserve University, VA Hospital, Perry Point, Maryland, and the University of Maryland, who have made this study possible.

Rationale

The development of the inventory has been guided by a psychological rationale integrated from a number of divergent fields: psychology, psychiatry, test theory, and sociology. The aim of this formulation is to provide a theoretical framework for using and interpreting the inventory, and to present a means of extending and clarifying its construct validity. The following assumptions summarize this tentative rationale:

1. The choice of an occupation is an expressive act which reflects the person's motivation, knowledge, personality, and ability. Occupations represent a way of life, an environment rather than a set of isolated work functions or skills. To work as a carpenter means not only to use tools but also to have a certain status, community role, and a special pattern of living. In this sense, the choice of an occupational title represents several kinds of information: the *S*'s motivation, his knowledge of the occupation in question, his insight and understanding of himself, and his abilities. In short, item responses may be thought of as limited but useful expressive or projective protocols.

2. The interaction of the person and his environment creates a limited number of favorite methods for dealing with interpersonal and environmental problems. The various HVPI scales are assumed to measure some of these common or favorite methods of adjustment. Translated into scale terms, peaks reveal the person's favorite methods whereas low points indicate the rejected methods of adjustment. Or, peaks may represent desirable roles and situations, and low points, threatening or distasteful roles and situations.

The foregoing assumption is predicated by another assumption; namely, that the various classes or occupational groups furnish different kinds of gratifications or satisfactions and require different abilities, identifications, values, and attitudes. This particular hy-

pothesis has extensive empirical support from studies relating "vocational interests" to personality variables, psychiatric status, values and attitudes. Typical studies in this area include those of Berdie (1), Darley (2), Forer (3), Garman (4), Gough (5), Laurent (7), Sternberg (11), Terman (13), and Weir (15).

3. The development of adequate adjustive techniques requires accurate discrimination among potential environments. The ability to discriminate potentially satisfying and beneficial environments from potentially dissatisfying and unhealthy environments is imperative for creative health. In this sense, the inventory is a miniature performance test of the *S*'s understanding of his surroundings in relation to himself; that is, the choice of an occupational title is a measure of the *S*'s insight and understanding as well as a sign of his motivation and his intellectual comprehension of the occupation in question. This third assumption has two corollaries which seem useful for interpreting "responsiveness" to the inventory.

3a. The total number of preferred occupations is a function of a number of personality variables. These hypothesized variables are supported by some of the evidence for the validity of the inventory, and by several scattered studies reported by Berdie (1), Weir (15), and others.

Specifically, the total number of preferred occupations is a function of dependency, aggressiveness, mood, degree of cultural introception, self-control, sociability, and defensiveness. Over-responsiveness suggests a lack of adequate discrimination which may be reflected in dependence, aggression, euphoria, over-intraception of the culture, impulsivity, sociability, frankness. In contrast, under-responsiveness appears indicative of greater independence, passivity, depression, rejection of the culture, over-control, withdrawal, and defensiveness.

3b. The inability to make discriminations among occupations is indicative of conflict and disorganized self-understanding. Just as the inability to make everyday decisions is a result of conflicting motivations, so the inability to make positive or negative choices of occupations (environments) within the inventory is a sign of conflict. In this sense, conflict is defined as divergent, inaccurate, or irreconcilable views about one's abilities, needs, and sources of gratification; and is accompanied by the chronic emotional upset which results from such conflict. In test terms, inability to make choices is reflected in the total number of omitted items.

4. Interest inventories are personality inventories. Interest and personality inventories are identical in principle and provide similar information about the person, although their content is quite diverse. Both kinds of inventories reveal how the *S* perceives himself and his milieu.

Construction of the Inventory

Preliminary Forms

The first form of the inventory was established by constructing eight a priori scales: Physical Activity, Intellectuality, Responsibility, Conformity, Verbal Activity, Emotionality, Reality Orientation, and Acquiescence. These scales were devised by reviewing the interest and vocational choice literature with reference to personality factors, and by using this information to create personality formulations for each scale. With the formulations as a guide, the author constructed the scales by selecting occupational titles which presumably epitomized the scale rationale. For example, the Physical Activity scale consists of occupational titles which imply motoric activities and skills. Typical titles in this scale include machinist, North Woods guide, forest ranger, electronic technician, etc. Scales are scored by counting the number of "preferred" occupations in a given scale.

The second form consists of a revision of the first form by means of an internal consistency analysis of the individual scales and of six added scales, forming a 14-variable test. The additional scales were Control, Aggressiveness, Mf, Status, Heterosexuality, and Infrequency. Next, all scales were intercorrelated and a cluster analysis was performed in order to eliminate scales and clarify scale meanings. This analysis suggested that 10 scales were sufficiently independent to merit retention and further revision.

Present Form

The third form was secured by performing an internal consistency analysis for those 10 scales which survived the cluster analysis of the second form. This analysis was accomplished with two samples of 300 male and female college freshmen. The method of analysis was to compare the upper and lower 25 per cent of each sample on each scale. The most discriminating items were then selected for the final scales, using the "Kelly Technique" nomographed by Lawshe (8).

The Masculinity-Femininity, Status, Infrequency, and Acquiescence Scales are exceptions to the construction method outlined above. These scales have a direct empirical development. The Status Scale was devised by reviewing three studies of occupational status which used the rank order method (6, 9, 10). Scale items were derived by making all items satisfy the data of all three studies. For example, "Physician" occurs as a high status choice in each study. The final scale was then keyed so that preferences for occupations of high status and aversions to occupations of low status result in a high status score.

Construction of the Mf Scale began with a comparison of the responses of 170 male and female high school students on the second form. This scale was then revised by a comparison of two groups of 150 male and female college freshmen.

The Infrequency Scale consists of 25 items which are rarely "liked" and 25 items which are rarely "disliked." It was constructed from frequency counts of the 300 test items for samples of 119 male and 100 female college freshmen. The 25 most popular and 25 least popular items were selected from each sample to form scales by reversing the scoring; that is, popular items are scored "dislike" and unpopular items are scored "like." Accordingly, high scores represent unusual or unpopular choices, whereas low scores represent popular choices.

The Acquiescence Scale provides an estimate of the *S*'s acquiescence, his tendency to respond "Appeal" or "Like." It is scored by counting the number of "Appeal" responses for Items 1 through 30. Although a more accurate estimate of acquiescence can be obtained by counting the "Appeal" responses for all test items, Items 1–30 provide a useful and economical estimate since Items 1–30 correlate .74 with Items 31–300 for males and .69 for females.

The third and present form of the inventory contains 13 scales in all: 3 response-set scales and 10 personality scales. The response-set

Table 1

Scale Matrix for Male and Female College Freshmen

	Scales											
	PA[a]	In	Re	Cf	VA	Em	Co	Ag	Mf	St	Inf	Ac
In	43		37	18	27	25	19	56	38	27	15	44
Re	05	16		11	15	01	52	40	01	51	05	64
Cf	07	05	41		38	13	01	32	−12	07	03	36
VA	35	05	52	83		47	12	41	05	32	−44	63
Em	36	70	37	24	27		04	19	29	06	−40	45
Co	−41	−39	01	01	10	−26		−02	−27	24	05	30
Ag	69	26	41	62	73	39	−09		11	52	42	12
Mf	53	20	41	30	55	30	09	66		30	02	17
St	−02	−17	26	29	50	−05	72	07	−49		−17	45
Inf	−60	−20	15	25	55	07	39	64	−54	11		37
Ac	22	65	66	66	55	54	−34	59	37	12	−58	

Note.—r_t of .30 is significant at .05 level. Intercorrelations for males are shown in lower left triangle of matrix; for females, in upper right triangle.

[a] Since the Physical Activity items for females are homogeneous with the Intellectuality items, they are included in the latter scale; consequently, the Physical Activity scale is omitted in the female form of the inventory.

scales include the: 1. Question Scale, or the number of omitted items, 2. Infrequency Scale, and 3. Acquiescence Scale. The remaining scales are: 1. Physical Activity, 2. Intellectuality, 3. Responsibility, 4. Conformity, 5. Verbal Activity, 6. Emotionality, 7. Control, 8. Aggressiveness, 9. Masculinity-Femininity, and 10. Status.

To secure two equivalent parts or tests, each scale has been halved and the halves separated to form two tests. This procedure establishes a useful format for determining reliability and presents an opportunity for determining the value of a short form. Table 1 shows the scale interrelationships for college male and female samples of 100 each. For males, the intercorrelations range from −.60 to .83, and average .36. For females, *r*'s range from −.44 to .64 and average .26.

Age and Intelligence [2]

The relation of HVPI scales to age and intelligence has been tested, since it is desirable for interpretative purposes to have the inventory relatively independent of these variables. Heterogeneous samples of high school and college students and of employed adults were combined to increase the age range in correlating age with HVPI scales. Males range in age from 15 to 77; females range from 17 to 60 years. In general, these relationships are either insignificant or of a low order. The correlation of −.64 between Control scale and age for females is the exception to these findings.

The relation between intelligence and HVPI scales was estimated by correlating the Wonderlic Personnel Test, a revision of the Army Alpha, with the HVPI scales; samples of employed hospital attendants were used. None of the resulting correlations are significant; however, the range of Wonderlic scores is somewhat attenuated so that significant relationships may occur in more heterogeneous samples.

[2] Tables summarizing the relation of the inventory scales to age and intelligence, and scale reliabilities have been deposited with the American Documentation Institute. Order Document No. 5689, remitting $1.25 for 35-mm microfilm or $1.25 for 6 by 8 in. photocopies.

Reliability [2]

The internal consistency of the third revision was estimated by correlating Part I against Part II of each scale, using samples of 100 male and female college freshmen. The scale divisions are assumed to be equivalent, since items were assigned to Part I or II randomly. For males, these coefficients corrected for length range from .72 to .95, with a median of .85. For females, they range from .68 to .90, with a median of .79.

An estimate of retest reliability was obtained from data supplied by Walsh (**14**), who tested a sample of 38 TB patients before and after lung surgery. The time interval between testings averaged about four months. The retest correlations range from .70 to .87, with a median of .75.

Validity

The inventory differentiates significantly a number of defined criterion groups which represent a wide range of personality differences. In general, the obtained differences support the construct validity of the inventory. These findings are summarized in the following sections.

Table 2

Mean Scale Differences for Matched Controls and Psychiatric Patients (Age, Occupational Class, and Status)

	Controls (N = 100)		Patients (N = 100)		
Scale	X	*SD*	X	*SD*	*P*
Inf*	13.2	5.9	18.2	7.4	<.001
Ac	12.5	5.8	10.4	6.9	<.05
PA	15.0	7.9	10.9	8.7	<.01
Int	10.5	7.6	8.7	7.8	<.10
Re	9.5	6.2	7.6	7.3	<.06
Cf*	10.4	6.7	7.5	8.1	<.01
VA	10.1	6.6	7.3	7.7	<.01
Em	8.1	7.9	8.2	8.2	—
Co*	11.6	7.2	13.3	5.8	<.07
Ag	13.3	8.4	10.0	8.9	<.01
Mf	12.3	4.4	9.9	4.0	<.001
St	10.1	4.5	9.9	5.0	—

* Variances significantly different beyond .05 level.

Table 3

Distributions of Profile Peaks for Controls and Psychiatric Patients

Scale	Controls ($N = 183$)	Patients ($N = 186$)
PA	38	23
Int	5	8
Re	19	17
Cf	41	27
VA	1	1
Em	12	20
Co	35	67
Ag	12	6
Mf	19	8
St	1	9

Note.—For this comparison, the control and patient samples have been increased by adding unmatched *Ss* to these groups in order to increase the "expected" cell frequencies. These additional *Ss* closely approximate the matched samples with respect to age, status, and occupation.

Controls and Psychiatric Patients

For this comparison, male samples of 100 controls and 100 psychiatric patients were matched for age, socioeconomic status, and principal occupation. The group distributions for these matched variables are not significantly different. The control group consisted of employed and unemployed adults tested in the process of employee selection and similar personnel studies at a VA psychiatric hospital. The patient group included 74 psychotic and 26 nonpsychotic patients, all without organic involvement. Most of them were tested near the end of their hospitalization, although a few were tested shortly after admission. Table 2 summarizes the results of the comparison.

The significant differences in Table 2 are in the expected directions; that is, a literal interpretation of the scale meanings yields descriptions for normals and for psychiatric patients which are congruent with stereotypes of normality and psychiatric status. These differences reveal that normals give common responses (Inf), are more responsive (Ac), are physically active (PA), conforming (Cf), verbal (VA), aggressive (Ag), and masculine (Mf). In contrast, patients give infrequent or unusual responses (Inf), are less responsive (Ac), less active (PA), nonconforming (Cf), less verbal (VA), more passive (Ag), and more feminine (Mf).

Similar differentiation is obtained between control and patient samples when distributions of high point scales are formed. For this comparison, individual profiles were classified by the highest scale in the profile. The resulting distributions for patients and con-

Table 4

Mean Scale Scores for Controls, Psychiatric Patients, TB Patients, and Psychopaths

Scale	Control ($N = 100$)	Psychopath ($N = 50$)	TB ($N = 61$)	Psychiatric ($N = 100$)	F
?	25.2	51.5	16.2	30.9	—
Inf	13.2	14.5	18.9	18.2	13.47***
Ac	12.5	10.3	11.3	10.4	2.02
PA	15.0	12.0	13.0	10.9	3.88**
Int	10.5	8.9	8.9	8.7	.96
Re	9.5	6.9	7.0	7.6	3.03*
Cf	10.4	6.9	9.3	7.5	4.16**
VA	10.1	7.7	8.1	7.3	2.62
Em	8.1	9.1	8.5	8.2	.18
Co	11.6	11.2	13.0	13.3	2.09
Ag	13.3	9.3	10.6	10.0	3.29*
Mf	12.3	10.4	10.9	9.9	5.57***
St	10.1	10.0	9.9	9.9	.04

Note.—Controls and psychiatric patients were matched for age, status, and occupation. Psychopaths and TB patients approximate first two groups for these variables. Note insignificant differences on status scale.

* p .05.
** p .01.
*** p .001.

Table 5

Significant Mean Score Differences

Scale	Criterion Groups	Sig Level
Infrequency	NP, TB > C, Pd	<.001, <.01
Physical Activity	C > NP, Pd	<.01, <.05
Responsibility	C > Pd, TB	<.05, <.05
Conformity	C > Pd, NP	<.01, <.01
Mf	C > Pd, NP, TB	<.05, <.001, <.05
Aggressiveness	C > Pd, NP	<.01, <.01

Table 6

Distributions of Profile Peaks for University of Maryland Colleges Entering Male Freshmen (N = 382)

Scales	Arts and Science (N = 101)	Engineering (N = 120)	Business (N = 96)	Education (N = 38)	Agriculture (N = 27)
Physical Activity	15	33	3	11	11
Intellectuality	25	39	7	2	9
Responsibility	22	5	18	13	1
Conformity	9	13	41	6	4
Verbal Activity	10	8	15	2	1
Emotionality	20	22	12	4	1

Entering Female Freshmen (N = 119)

Scales	Arts and Science (N = 59)	Home Economics (N = 46)	Business (N = 17)	Education (N = 58)	Nursing (N = 19)
Intellectuality	14	4	3	5	7
Responsibility	10	4	1	20	8
Conformity	14	12	10	18	3
Verbal Activity	12	11	2	10	0
Emotionality	9	15	1	5	1

Note.—For males, distributions of scale peaks for Arts and Sciences, Engineering, and Business are significantly different at the .01 and .001 levels, respectively. For females, distribution for Arts and Sciences, Home Economics, and Education are significantly different at the .10, .06, and .01 levels, respectively. For this comparison, only the first six inventory scales were employed.

trols, shown in Table 3, are significant beyond the .001 level.

Normals, Psychopaths, Tuberculosis, and Psychiatric Patients

The study reported above was extended by comparing the matched control and psychiatric samples with two additional samples of 61 TB patients and 50 criminal psychopaths.[8] These samples are assumed to have a crude comparability with respect to age, socioeconomic status, and intelligence. The mean ages for the four groups range from 27.0 to 33.2 years; the four means on the status scale range only from 9.9 to 10.1 raw score points; there is a similar distribution of principal occupations in each of the four groups.

[8] The prison psychopaths were selected by Lloyd B. Calkins, psychologist for the Maryland penitentiary. All *Ss* in this sample were diagnosed by a test battery including, at a minimum, the MMPI, W-B, and DAP.

The group means were tested first by a simple analysis of variance; then individual *t* tests were performed for the significant analysis of variance tests. The results of these analyses are shown in Tables 4 and 5.

College Choice

An additional study attempted to differentiate the profiles of university freshmen on the basis of their college choice. The sample included groups of male freshmen majoring in business, arts and sciences, and engineering. For the analysis, 209 profiles were classified by high point scales among the first six scales of the second revision and frequency distributions formed for each college. χ^2 tests revealed that differences between any pair of colleges are significant beyond the .05 level.

The following year, this test was replicated with a sample of 317 freshmen, using the third revision of the inventory. Two of the differences among scale peak distributions are significant at the .001 level and a third at the .01 level. In addition, these differences are substantial. The engineers, for example, have 60 per cent of their peaks among the Physical Activity and Intellectuality scales, whereas the business freshmen have only 10.4 per cent of their peaks among these scales. Likewise, the business freshmen have 77 per cent of their peaks among the Responsibility, Conformity, and Verbal Activity scales, whereas the engineers have 22 per cent of their peaks among these scales. Table 6 summarizes this evidence along with several small samples not amenable to statistical test.

Received January 15, 1958.

References

1. Berdie, R. F. Likes, dislikes, and vocational interests. *J. appl. Psychol.*, 1943, **27**, 180–189.
2. Darley, J. G. A preliminary study of relations between attitude, adjustment, and vocational interest tests. *J. educ. Psychol.*, 1938, **29**, 467–473.
3. Forer, B. R. Personality dynamics and occupational choice. *Amer. Psychologist*, 1951, **6**, 378–379. (Abstract)
4. Garman, G. D. The Strong Vocational Interest Blank as a measure of manifest anxiety. *Amer. Psychologist*, 1954, **9**, 372–373. (Abstract)
5. Gough, H. G., McKee, M. G., & Yandell, R. J. Adjective check list analyses of a number of selected psychometric and assessment variables. Berkeley: Inst. Pers. Assessment and Research, Univer. Calif., 1955. (Mimeo.)
6. Hall, C. W. The prestige values assigned to a group of 249 occupations by 200 independent raters. Western Reserve Univer. (Mimeo.)
7. Laurent, H. A study of the development backgrounds of men to determine by means of the biographical information blank the relationship between factors in their early backgrounds and their choice of professions. Unpublished doctoral dissertation, 1951, Western Reserve Univer.
8. Lawshe, C. H., Jr. A nomograph for estimating the validity of test items. *J. appl. Psychol.*, 1942, **26**, 846–849.
9. National Opinion Research Center. Jobs and occupations: A popular evaluation. *Opinion News*, 1947, **9**, 3–13.
10. Randall, H. Occupational Prestige. Occupational Planning Committee, Cleveland Welfare Federation, 1951. (Mimeo.)
11. Sternberg, C. Personality trait patterns of college students majoring in different fields. *Psychol. Monogr.*, 1955, **69**, No. 18 (Whole No. 403).
12. Strong, E. K., Jr. Vocational interests of men and women. Stanford: Stanford Univer. Press, 1943.
13. Terman, L. M. Scientists and nonscientists in a group of 800 gifted men. *Psychol. Monogr.*, 1954, **68** (Whole No. 7).
14. Walsh, R. P. Vocational interests, their stability and personality correlates in hospitalized tuberculosis patients. Unpublished master's thesis, Univer. Maryland, 1956.
15. Weir, J. R. An attempt to identify vocational interest profiles within a neuropsychiatric population. Unpublished doctoral dissertation, Univer. California, 1951.

Journal of Applied Psychology
Vol. 42, No. 5, 1958

Length of Work Periods in Visual Research [1]

Miles A. Tinker

University of Minnesota

The duration of the task employed in investigating visual efficiency has varied considerably from one reported experiment to another. In general the tendency has been to use relatively short work periods, i.e., one to five minutes. The reliability and validity of measurement as related to length of work period is an important issue. In an earlier study by Tinker (3), it seemed that a prolonged reading task, such as 10 minutes, was necessary to achieve valid results in certain critical comparisons where differences tend to be small. One of the difficulties in this earlier study (3) was that results for brief work periods in a 1928 investigation were compared with prolonged work periods in the 1955 report. It now seems obvious that such comparisons should be made for work periods by the same group of subjects.

In the 1955 study, Tinker (3) did show that the results were the same in comparing italic with roman lower case printing in each of three successive 10-minute work periods and in a 30-minute work period. Similar trends were found in comparing lower case with all-capital printing in each of four successive four-minute work periods and in a 16-minute work period. The italic was read significantly slower than the roman and the all-capital material was read significantly slower than lower case printing. Reliability of measurements was fairly high: a mean of .84 for one and .87 for the other study. These results, therefore, indicate that 10 minutes is as satisfactory as 30 minutes in the former comparison and that four minutes is as satisfactory as 16 minutes in the latter comparison. It is now desirable to explore results obtained for short work periods, i.e., from 1½ to 10 minutes.

The present experiment was designed to investigate the use of relatively short work periods in one type of visual research. It appears that the length of work period that needs exploring ranges from 1½ to 10 minutes. In the study (3) cited above it is clear that 10 minutes was as satisfactory as longer periods in one comparison (italics vs. roman), and, similarly, 4 minutes was satisfactory in the other (all-capitals vs. lower case). The present experiment is concerned with speed of perception in reading under various levels of illumination.

Materials and Procedure

Tinker's Speed of Reading Test (2) was employed. Since comprehension is constant, the test measures virtually speed of perception in reading as a single variable. The two forms of the test are approximately equivalent. Both were set in regular (roman) 10-point Excelsior type face with 2-point leading in a 20-pica line width on eggshell paper stock. The 180 university sophomore *Ss* were tested individually in a light laboratory which provided well-dispersed indirect illumination. In Group A, the Control Group, Forms I and II of the test were given under 25 foot-candles of light. In Group B Form I was given under 5 foot-candles and Form II under 25 foot-candles of light. And in Group C Form I was administered under 25 and Form II under 200 foot-candles of light. The time limit was 10 minutes for each form, with marks on the test to indicate the end of 1½ and 5 minutes of work. Successive *S*'s for Groups A, B, and C were systematically rotated, i.e., *S* No. 1 was put in Group A, No. 2 in Group B, No. 3 in Group C, No. 4 in Group A, etc.

Results and Discussion

The results are given in Table 1. Inspection of the table reveals the following trends: (*a*) The reliability of measurement is high, varying from .82 to .97 with a median coefficient of .95. (*b*) In Group B the results show that similar trends occurred with 1½, 5, and 10 minutes of reading, i.e., there was no significant difference in performance under 5 vs. 25 foot-candles of light. (*c*) In Group C there was no significant difference in performance under 25 vs. 200 foot-candles of light with a work period of 1½ minutes. With the 5-minute and 10-minute work periods, the read-

[1] The writer is grateful to the University of Minnesota Graduate School for a research grant to finance this study.

Table 1

Effect of Illumination on Speed of Reading With Varying Lengths of Reading Time (N = 60 in each test group, 180 in all)

Test Group	Time Limit	Test Form and Foot-candles	Mean	*SD*	Diff. Between Means in Paragraphs[a]	Diff. Between Means in Per Cent	*r*	Diff. / S.E. Diff.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A	First	I, 25	13.62	4.05	0.0	0.0	.83	0.00
	1½ min.	II, 25	14.12	4.02				
A	First	I, 25	46.77	11.28	0.0	0.0	.95	0.00
	5 min.	II, 25	46.62	12.09				
A	First	I, 25	94.97	22.97	0.0	0.0	.97	0.00
	10 min.	II, 25	95.42	26.36				
B	First	I, 5	12.87	3.21	+0.25	+1.9	.82	1.01
	1½ min.	II, 25	13.62	3.19				
B	First	I, 5	44.27	9.58	+0.18	+0.4	.92	0.37
	5 min.	II, 25	44.30	9.99				
B	First	I, 5	89.80	20.58	−0.05	−0.1	.95	0.06
	10 min.	II, 25	90.20	21.58				
C	First	I, 25	14.40	3.74	−0.10	−0.7	.86	0.35
	1½ min.	II, 200	14.80	4.34				
C	First	I, 25	48.92	10.57	−1.40	−2.9	.95	2.55
	5 min.	II, 200	47.37	12.49				
C	First	I, 25	99.63	22.71	−3.07	−3.1	.97	3.55
	10 min.	II, 200	97.02	25.60				

[a] The differences in Column 6 are "corrected" by the amount of the difference between the mean scores of Form I and Form II in Test Group A which serves as a control group. The "corrections" amount to −0.50 for the first 1½ minutes, +0.15 for the first 5 minutes, and −0.45 for the 10 minutes. Original computations were carried to 4 decimal places.

ing was retarded significantly under the 200 foot-candles of light, at the 2 and 1 per cent levels, respectively.

The results presented here suggest that 1½ minutes of reading is just as satisfactory as the longer periods of 5 and 10 minutes for 5 vs. 25 foot-candles illumination. It might be noted that Paterson and Tinker (1), page 178, found similar trends in typography studies, i.e., 1¾, 5¾, and 10 minutes working time all yielded comparable results. For 25 vs. 200 foot-candles the trend is different as noted above.

The fact that speed of visual perception in reading decreased under 25 vs. 200 foot-candles for work periods of 5 and 10 minutes needs some interpretation. It is suggested here that possibly glare was operating to reduce efficiency under the 200 footcandles of light. It is possible that the 1½-minute work period was too short for the glare to decrease efficiency but that the glare did become effective under the longer work periods of 5 and 10 minutes. Note in Table 1 that the decrease in efficiency was greater for the 10 than for the 5 minutes. Similar trends have appeared for high light intensities in another study not yet published.

Summary and Conclusions

1. The purpose of this experiment is to provide additional information on the effect of length of work periods in visual research.

2. Work periods of 1½, 5, and 10 minutes were used while investigating the relative efficiency of speed of perception in reading under 5, 25, and 200 foot-candles of light.

3. There were 180 university sophomore *Ss* who were tested individually under well-controlled indirect illumination.

4. The three work periods yielded the same results for the 5 vs. 25 foot-candles comparison, i.e., there was no significant difference in performance with change in illumination. With the 25 vs. 200 foot-candles comparison, there was no significant difference for the 1½-minute work period but speed of reading was significantly slower under the 200 foot-candles for both the 5- and 10-minute work periods. It is suggested that glare was having a deleterious effect under the high illumination for the longer work periods.

5. One may tentatively conclude from the above results that work periods as short as 1½ minutes may be safely used in studying the relation between illumination and visual efficiency.

Received January 17, 1958.

References

1. Paterson, D. G., & Tinker, M. A. *How to make type readable.* New York: Harper, 1940.
2. Tinker, M. A. *Tinker speed of reading test.* Minneapolis: University of Minnesota Press, 1955.
3. Tinker, M. A. Prolonged reading tasks in visual research. *J. appl. Psychol.*, 1955, **39**, 444–446.

Journal of Applied Psychology
Vol. 42, No. 5, 1958

Communication Restraints, Group Flexibility, and Group Confidence[1, 2]

Robert C. Ziller

Fels Group Dynamics Center, University of Delaware

Interpretations of the results of experiments concerned with group problem-solving productivity are necessarily complicated by qualifications involving unique task demands (7). In an attempt to avoid this limitation, the study reported here concerns the correlates of a postulated fundamental dimension of group problem-solving behavior, group flexibility (5). More explicitly, the present study explores the relationship between selected group characteristics and group flexibility operationally defined as the ability of a group to reorganize to meet the time demands of a new situation. In addition, an investigation of group confidence as an affective concomitant of group performance was undertaken to further the understanding of group problem-solving behavior.

Problem

In general, it was proposed that the more flexible groups are those with more completely open communication systems or groups with less communication restraints.

The following characteristics of the leaders and groups were selected as independent variables, the F-scale score, the conformity of the leader, and the attraction of the group.[3]

Leaders with high F-scale scores were expected to structure the communication network of the group around themselves as the focal person, thereby restricting information exchange and group flexibility.

An attractive group is described simply as one in which the group members feel secure, feel as if they "belong," know what to expect of others and can communicate with them without restraint (3). There is evidence (6) that members of highly attractive groups tend to deviate from the group norms when the norm is discordant with reality. In such groups, the members are more disposed to critically evaluate established procedures in the light of new circumstances and evolve an organization more in accord with the demands of the given situation.

Finally, it was anticipated that the extent to which the leader is willing to modify his judgment toward the group norm (conformity) is related to group flexibility through the intervening variables of the relative openness of the group's communication network. It was proposed, provisionally, that groups whose leaders are moderately willing to modify their judgment toward the group norm are best prepared to adapt to the requirements of a new situation.

Experimental Procedure

Subjects

Altogether, 96 B-29 or B-50 aircrews comprising about 1,000 men were involved in the experiment. For the most part each crew was composed of 11 men of whom five were officers (including the assigned aircraft commander or leader) and six were airmen. However, some of the crews were as small as eight men and others as large as 15. All subjects had been members of the same crew for a minimum of three months. Some members had been together for as much as a year.

Procedure

Immediately preceding the problem-solving sessions, the members of a group completed sociometric

[1] This paper is a condensation of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Michigan, 1955. Special acknowledgment must be accorded W. Clark Trow, Alvin Zander, Leonard Berkowitz, and E. Paul Torrance for their insightful critical reading of the original manuscript.

[2] This research was carried out under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

[3] The selection of the variables was also influenced, in part, by time considerations, since the *Ss* were available for a maximum of one hour.

and group attraction questionnaires. Then they seated themselves around a rectangular area and the general nature of the group task was described. The directions read, in part, were:

"The object of the test is to make as high a crew score as possible; that is, the highest score that your crew can produce using any means at your disposal. The way you go about this is entirely up to you. You may hand in one answer sheet for the crew, an answer sheet for each crew member, or any other arrangements you wish to make. When two or more answer sheets contain answers to the same question, they will be averaged to provide a crew score for that question."

Any questions arising from the group were answered by rereading the directions. At this point a questionnaire purported to be a test of the individual's ability to estimate accurately the group's performance was administered. Actually, the questionnaire provided the data for the measurement of group confidence. Then the members were allowed one minute in which to examine a sample exercise, and then sixteen minutes in which to complete all eight exercises comprising the group task. During their performance, an observer recorded the group's method of approach to the task. Immediately following the problem-solving session, the crew members completed the F scale and the "conformity" scale.[4]

The Group Performance Test

In the experiment, each crew member received a copy of Test 401-B, one of the tests in the Intellectual Talents Battery being developed for use in the Air Force by the Personnel Research Laboratory, Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. The test booklet consisted of eight rather long exercises. In each exercise, a problem was raised and 15 facts presented that had a bearing on the solution to the problem. The *Ss* were instructed to choose five of the 15 facts most important in reaching a decision on the problem. The respondent selected the most important facts in deciding whether a certain enemy-held island base was being developed as an air base or a submarine base, what type of home a married officer should buy, what man a squadron commanding officer should recommend for Officer Candidate School, what city should be selected for a recruiting office, what site a manufacturer should choose for a factory, which of two devices to train radar observers should be purchased by the Air Force, which secretary should be hired, and which of two types of visual aids should be used by Air Force schools.

Because of the complexity of the decision-making situation, it was impossible for an individual or a discussion group to complete all eight problems efficiently within 16 minutes. The task was completed most expeditiously by splitting the group into two or more subgroups and apportioning the problems among them.[5]

Unpublished reports describing the development of the instrument indicate that measures of its interval consistency by the split-half technique were low ($r = .37$).

Measuring the Variables

F Scale

Two items were eliminated from the original F Scale when it was observed that Air Force personnel were openly derisive regarding the one, while the other appeared outdated. Reported reliability measures of the original instrument range from .81 to .91 with an average of .87 (1, p. 251). Scores were tabulated and divided into high and low categories by dividing the distribution at the median.

Conformity of the Leader

The measure of conformity included the following items:

Supose you found yourself holding an opinion which was not in agreement with an opinion held by the other people in your crew and this opinion was on an issue of importance both to you and to the crew. In each set of statements below, check the one alternative which would best express how you would feel and what you would do if such a situation existed in your crew.

1. I would feel: (check one) a. Not at all concerned about my differing from the crew opinion, b. Slightly concerned about my differing from the crew opinion, c. Moderately concerned about my differing from the crew opinion, d. Greatly concerned about my differing from the crew opinion.

2. If I found that my opinion differed from that of this crew, I would feel: (check one) a. Just as sure of my own opinion, b. Less sure of my own

[4] This was originally a four-item scale used by Alvin Zander, Research Center for Group Dynamics, University of Michigan, and was used in the present study at his suggestion.

[5] The methods used by the crews in the present experiment were anticipated through observations of aircrew performance on similar tasks (13).

opinion, c. Much less sure of my own opinion, d. Very much less sure of my own opinion.

3. If I found that my opinion differed from that of this crew, I would have: (check one) a. A desire to maintain my own opinion, b. Hardly any desire to change my own opinion to agree with that of the crew, c. A moderate desire to change my opinion to agree with that of the crew, d. A strong desire to change my opinion to agree with that of the crew, e. A very strong desire to change my opinion to agree with that of the crew.

In calculating the total score, the alternatives of each question were assigned a weight of one to four or one to five, in accordance with the number of options. The distribution of scores was divided into high, medium, and low categories in accordance with the considerations described initially. The low range included scores 10, 11, 12, and 13. The cut-off scores were dictated by the necessity of a workable number of cases in the middle range.

Group Attraction

The index of group attraction followed directly from Festinger's definition of group cohesiveness; that is, the resultant of all forces acting on the members to remain in the group (2, p. 164). The term "group attraction" was used rather than the more general term "cohesiveness" since the two questions involved in this index did not measure the forces acting to prevent the individual from leaving the group.

The index was derived from the following questions:

1. If one man from each crew were ordered to remain at the home base, rather than to attend Survival School and you were the man chosen, how would you feel about this? (check one) a. I'd feel very lucky, b. I wouldn't mind, c. It wouldn't matter one way or another, d. I wouldn't like it, but would not feel too strongly about it, e. I wouldn't like it a bit.

2. How would you feel about being transferred to another crew or about your crew being split up as replacements for other crews?

The alternatives were the same in each of the items.

In calculating the index of group attraction, the alternatives of each item were weighted from one to five, respectively. The score of each individual group member was obtained by a simple summation of the weighted responses. The mean score of the group provided the desired index. An index of the reliability of the measure was calculated using the split-half technique in which each crew was divided into two subgroups with approximately the same number of officers and airmen. A Pearsonian correlation of .46 was obtained by correlating the average scores of the 96 pairs of subgroups.

Table 1

Analysis of Variance of the Group Approach to the Problem and Group Score

Group Approach	*N*	Average Score
a	37	8.15
b	20	10.82
c	13	12.39
d	8	14.13
e	18	14.98

Note.—$F = 18.03$; $df = 4$ and 92; $p > .01$.

Group Flexibility

An operational definition of group flexibility was derived from the five methods of approach to the task: (*a*) group members worked individually with little or no interaction; (*b*) group worked as a unit discussing each problem in turn; (*c*) group began working as a unit (as in Method *b*) but, recognizing that there was insufficient time to complete the task by their initial method, changed to Method *e*; (*d*) group worked as a highly organized unit, limited the time spent on each problem to two minutes and appointed a time keeper and recorder; (*e*) group subdivided into two, three, four, or five sections and divided the problems accordingly.

The frequencies with which each method was employed are given in Table 1.

While the study is concerned primarily with process rather than product variables and the low reliability of the instrument notwithstanding, it was observed that the method of approach placed definite limits on the score attained.[6] This observation was easily tested

[6] Correct answers to the items were provided by the test writers. In cases where one answer sheet was submitted by the group, the group scores were obtained from these single sheets by applying a simple correction formula ($R - \frac{1}{4}W$). When members of the same groups submitted answer sheets which overlapped, the mean of the scores became the group score.

through a simple analysis of variance between method and group score (see Table 1). It was also apparent that the mean score increased in direct relation to the method employed. The highest scores were attained by groups which used the subgroup or highly-organized group approach.

Groups using the approaches which reflected an awareness of the need to reorganize to meet the time demands of the problem, Methods *c*, *d*, and *e*, were compared with groups using approaches *a* and *b*. Reorganization is used here in the sense that the aircrew was already organized as a highly integrated team with a single leader and was accustomed to performing as a single unit. Flexible crews were those who were able to overcome this organizational set and reorganize in accordance with the time demands of the new situation. Actually however, two inappropriate sets seemed to have been involved: the usual group-test-situation set in which comparisons of answers are not permitted (Method *a*) and the set that crews must perform as a single free-discussion unit (Method *b*).

Group Confidence

The individual group members indicated their degree of confidence in the group by responding to the following question:

How well do you think your crew will do in this problem-solving situation?

—— superior: better than 90% of the other crews.
—— above average: better than 70% of the other crews.
—— average: about as well as 50% of the other crews.
—— below average: only as well as 30% of the other crews.
—— low: only as well as 10% of the other crews.

The alternatives of the question were weighted from 1 to 5, with the lower score indicating the greater confidence. The split-half estimate of reliability was .55 ($p = .001$).

The relationships among the three independent variables were calculated (Table 2). All correlations were low. Along with the moderate reliability estimates of the measures, the evidence permits the conclusion that the experiment involves three relatively unrelated predictors. It was interesting to note, however, that groups led by the more conforming leaders tended to be more attractive.

Table 2

Intercorrelations Among the Three Independent Variables

	I	II	III
I Rigidity of the leader		−.12	−.18
II Conformity of the leader			.29*
III Group Attraction			

* Significant at the .01 level.

Results

The data did not permit a pattern analysis involving the three independent variables simultaneously as this would have generated cells without cases. Therefore, chi-square tests were calculated from the 2×2 tables formed by tabulating the frequency with which groups, dichotomized according to the independent variables, used the flexible or non-adaptive approaches to the task.

The analysis of the data with regard to the relationship between the leader's F-scale score and the criterion of group flexibility is presented in Table 3. The differences with regard to the flexibility criterion were in the expected direction but statistically significant only at the 10 per cent level of confidence. Nevertheless, it was noted that 50 per cent of the *low* groups used the approaches which reflect flexibility as compared with 31 per cent of the ***high*** groups.

With regard to the association between leader conformity and group flexibility (Table

Table 3

The Relationship Between the Leaders' F-Scale Scores (Dichotomized) and Group Flexibility

	Group Approach				
F Scale	*a*	*b*	*c*	*d*	*e*
High	22	11	4	3	8
Low	15	8	9	4	10

Chi square (*a* and *b* vs. *c*, *d*, and *e*) = 3.43; $p = .10$

Note.—The total number of groups reported using each approach is not always consistent with the number in Table 1 due to incomplete data with regard to two crews.

Table 4

The Relationship Between the Leaders' Conformity Indices and Group Flexibility

Conformity	Group Approach				
	a	*b*	*c*	*d*	*e*
High	12	3	0	1	3
Medium	11	13	12	2	12
Low	14	4	1	4	3

Chi square (*a* and *b* vs. *c*, *d*, and *e*) = 6.33; $p = .05$

Note.—The total number of groups reported using Approach *d* is not consistent with the number reported in Table 1 due to incomplete data.
High and low conformity frequencies were pooled throughout.

4), it was seen that a greater percentage of groups in which the leader's C-scale scores indicate moderate conformity used approaches reflecting greater flexibility. The approaches representing a high degree of flexibility were used by 52 per cent of the moderate groups but only by 27 per cent of the *high-low* groups. Thus, it was found that groups in which the leader was moderately concerned about the group's opinions in comparison with those groups in which the leader was extremely concerned or unconcerned were more flexible in this problem-solving situation.

The results with regard to the relationship between group attraction and group flexibility were not statistically significant and no trend was in evidence (see Table 5).

It was apparent, however, that the measure of group flexibility employed was not sufficiently sensitive; far too many groups were required for analysis. It is entirely possible, however, that groups in a setting other than the military would avoid the approach to the task in which the group acted merely as an uncommunicative collection of individuals. In this event, the time required to shift to the subgroup approach to the task might serve as a superior measure of group flexibility. Of course, it still remains to be shown to what extent flexibility, as defined in the present study, is generalizable.

Table 5

The Relationship Between Group Attraction and Group Flexibility

Group Attraction	Group Approach				
	a	*b*	*c*	*d*	*e*
High	18	9	11	1	9
Low	19	10	2	7	9

Chi square (*a* and *b* vs. *c*, *d*, and *e*) = .29; $p = .70$

Table 6

Analysis of Variance of Group Confidence According to Leaders' F-Scale Scores and Group Attraction Indices

Source of Variance	*df*	Variance Estimate	*F*	*p*
Leader F Scale	1	16.62	1.88	
Group Attraction	1	149.56	16.90	.001
Interaction	1	35.88	4.05	.05
Groups	90	8.85		

Note.—When leaders' F-scale scores were categorized as high and low, and Group Attraction scores were similarly divided, the mean Group-Confidence indices for groups within the categories were (lower scores indicate greater confidence in the group):

Leader F-Scale Score	Group Attraction		Total
	High	Low	
High	1.80 ($N = 23$)	2.19 ($N = 25$)	2.00
Low	1.97 ($N = 25$)	2.09 ($N = 21$)	2.03
Total	1.89	2.14	2.01

Table 7

Analysis of Variance of Group Confidence According to Leaders' Conformity Scores and Group Attraction Indices

Source of Variance	*df*	Variance Estimate	*F*	*p*
Leader Conformity	2	23.88	2.66	
Group Attraction	1	91.60	10.20	.01
Interaction	2	2.91		
Groups	88	8.98		

Note.—When Leaders' Conformity scores were categorized as high, medium, and low, and Group Attraction indices were dichotomized, the mean Group Confidence indices for groups within the categories were (lower scores indicate greater confidence in the group):

Group Attraction	Leader Conformity			Total
	High	Medium	Low	
High	1.77 ($N = 12$)	1.93 ($N = 30$)	1.93 ($N = 6$)	1.89
Low	2.00 ($N = 6$)	2.10 ($N = 20$)	2.23 ($N = 20$)	2.14
Total	1.85	2.00	2.16	2.01

Table 8

Analysis of Variance of Group Confidence According to Leaders' F-Scale and Conformity Scale Scores

Source of Variance	*df*	Variance Estimate	*F*	*p*
Leader Conformity	2	56.52	6.03	.05
Leader F Scale	1	1.64		
Interaction	2	4.90		
Groups	88	9.37		

Note.—When leaders' conformity scores were categorized as high, medium, and low, and leaders' F-scale scores were dichotomized, the mean Group Confidence indices for groups within the categories were (lower scores indicate greater confidence in the group):

Leader F-Scale Score	Leader Conformity			
	High	Medium	Low	Total
High	1.65 ($N = 8$)	1.98 ($N = 24$)	2.21 ($N = 16$)	2.00
Low	2.00 ($N = 10$)	2.01 ($N = 26$)	2.09 ($N = 10$)	2.03
Total	1.84	2.00	2.16	2.01

In analyzing the data with regard to group confidence, factorial designs for cells of unequal size were employed (**11**, p. 285). (In Tables 6, 7, and 8 the entries in each cell are the mean group confidence scores for the groups falling in the category.) The smaller means indicate greater confidence.

The results with regard to group attraction and leader conformity were statistically significant and indicate that members of the more attractive groups whose leaders claim that they tend to conform to the opinion of the majority express greater confidence in their group's ability to succeed in a new situation.

The main effects of leader F-scale scores were not significant. However, when leader F-scale scores and group attraction indices were studied conjointly, interaction effects were significant (see Table 6). The greatest confidence was expressed by attractive groups, the leaders of which had high F-scale scores.

While group productivity was not central to the study, nevertheless the relation between the independent variables and the group scores was explored by means of analysis of variance designs similar to those in Tables 6, 7 and 8. As might have been anticipated on the basis of the low reliability index of the performance test, the results were not statistically significant nor were there any apparent trends.

Discussion

In terms of the communication framework sketched at the outset, both the more flexible groups and the more confident groups were generally found to be those with the more open communication systems or communication systems with relatively fewer restraints on the member's interaction. Thus, the results with reference to group flexibility suggest that groups with moderately conforming and *Low* F leaders adapted more readily to the requirements of the new situation. The more confident groups were those with highly conforming leaders and those to which the group was highly attractive for the members.

The apparent inconsistency in the findings regarding the conformity of the F leader was anticipated and interpreted as a qualification rather than as a directly contradictory result. That is, it was anticipated that while overly nonconforming leaders would tend to restrict communication among the members, the overly conforming leader, on the other hand, would tend to lack the confidence necessary to initiate group action. Thus, the characteristics of the leader which the group members find desirable—high conformity—may not be desirable with reference to other criteria—group flexibility.

The relationship between the independent variables presumed to be related to open communication systems and subsequently found to be related to group confidence corroborates the results of earlier experiments involving communication networks. In these earlier studies, it was revealed that, in general, members of groups with more open communication networks tend to have high group morale (**9**, **10**). However, it should be emphasized that the open communication systems in the present study are presumed to be psychologically determined rather than mechanically determined as was the case in the earlier studies.

With reference to the specific independent variables, the main effects of leader F-scale scores were not significant with regard to

group confidence. However, when leader F-scale scores and group attraction indices were analyzed conjointly, interaction effects were significant (see Table 7). The greatest confidence was expressed in attractive groups, the leaders of which had high F-scale scores; while the least confidence was expressed in low attraction groups, the leaders of which had high F-scale scores. If high group attraction is interpreted as indicating homogeneity of the members' F-scale scores, the results are in agreement with those presented in an earlier study relating homogeneity of the members' F-scale scores to group morale (4).

Contrary to a number of previous studies (1, 8, 12) the results reported above suggest a positive aspect of "authoritarianism." Presumably, the positive or negative aspects of authoritarianism with regard to the leader emerge only in conjunction with other leader and group characteristics and under particular task demands.

Summary and Conclusions

This study was designed to explore the relationship between selected group structure variables and the group's ability to adjust to the requirements of a new situation (group flexibility) and the group members' expressed confidence in the ability of the group to succeed in a problem-solving situation. The independent variables selected for analysis included the F-scale scores and "conformity" of the assigned leader and group attraction.

The *Ss* were 96 aircrews comprising about 1,000 men. The group task required the completion of an eight-item intelligence examination in a period of time which demanded a change in the group's customary operating procedure.

In the more flexible groups the leaders scored low on the F-scale and moderately high on a scale of conformity. It was also found that greater confidence in the group was expressed by members of high attraction groups and groups whose leaders tended to conform to the group members' opinions. In addition, interaction effects of leader F-scale scores and group attraction with regard to group confidence were statistically significant. It was concluded that groups with more open communication systems (group with fewer communication restraints) are more flexible and more confident.

Received January 27, 1958.

References

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harpers, 1950.
2. Festinger, L., Schacter, S., & Back, K. *Social pressures in informal groups: A study of a housing project*. New York: Harpers, 1950.
3. Hartley, E. L., & Hartley, R. E. *Fundamentals of social psychology*. New York: Knopf, 1952.
4. Haythorn, W. The effects of varying combinations of authoritarian and equalitarian leaders and followers. *J. abnorm. soc. Psychol.*, 1956, **53**, 210–219.
5. Hemphill, J. K. Situation factors in leadership. Ohio State Univer.: *Bureau of Educational Research Monographs*, **32**, 1950.
6. Kelley, H. H., & Shapiro, M. M. An experiment on conformity to group norms when conformity is detrimental to group achievement. *Amer. soc. Rev.*, 1954, **19**, 667–677.
7. Kelley, H. H., & Thibaut, J. W. Experimental studies of group problem-solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954.
8. Rokeach, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
9. Shaw, M. E. A comparison of two types of leadership in various communication nets. *J. abnorm. soc. Psychol.*, 1955, **50**, 127–134.
10. Shaw, M. E., & Rothschild, G. H. Some effects of prolonged experience in communication nets. *J. appl. Psychol.*, 1956, **40**, 281–286.
11. Snedecor, G. W. *Statistical methods*. (4th ed.) Ames, Iowa: Iowa State College Press, 1946.
12. Stern, G. G., Bloom, B. S., & Stein, M. I. *Assessment of personality II: Studies of the relationship between personality syndromes, learning situations, and learning outcomes*. Chicago: Univer. Chicago, 1952.
13. Torrance, E. P. Crew performance in a test situation as a predictor of field and combat effectiveness. Washington, D. C.: Human Factors Operations Research Laboratories, 1953. (HFORL Report No. 33.)

Journal of Applied Psychology
Vol. 42, No. 5, 1958

The Comparative Effectiveness of Some Psychological and Physiological Measures in Ranking the Impact of Diverse Environmental Conditions [1]

Bernard J. Fine

QM Research and Engineering Center Laboratories, Natick, Massachusetts

In the course of evaluating the thermal protection qualities of military clothing, it is frequently necessary to measure differences in the responses of soldiers along several physiological dimensions. These measurements are generally taken, of necessity, under controlled conditions, and are not readily adaptable to field use. Furthermore, they are both time consuming and expensive since they involve the use of fairly elaborate equipment both in obtaining the measures and in processing the data.

The present study investigates, in two parts, the use of subjective rating scales as substitutes for the more complex physiological measurements.

The subjective aspects of man's thermal environment have had little systematic study. Most of the work that has been done is psychophysical in nature and involves, primarily, the determination of thresholds of temperature sensation (4, 6, 7, 8). Several studies have investigated the relationships between personality variables and subjective responses to various environmental factors (1, 3, 10, 11). There have also been some attempts to get at the psychological and physiological bases of subjective reports of warmth or coldness (1, 2, 5, 11). Insofar as is known, there have been no studies comparing psychological with physiological measurements in their effectiveness in ranking a number of different environments according to the thermal state of individuals in those environments. Such a comparison is reported in the research described below.

[1] The author wishes to thank P. Iampietro for his cooperation in making the physiological data available and for his helpfulness with regard to the physiological aspects of the study.

Study Number One

Method

Subjects. The subjects (*Ss*) were six enlisted men stationed at the Quartermaster Research & Engineering Command as volunteer test subjects. They are described in greater detail in a paper by Iampietro et al. (9). They were considered to be "normal" on the basis of psychological and physiological measurements.

Procedure. The six *Ss*, clad only in shorts, were exposed as a group to different environmental conditions on eight successive days (excluding one weekend). The conditions represented combinations of ambient temperatures of 50 and 60 degrees F. with humidities of 30 and 95 per cent and wind speeds of zero and 10 miles per hour. The order in which the conditions were presented was randomly determined. Each exposure lasted two hours each day. A one-hour "control" period preceded each exposure and a one-half hour "recovery" period followed each exposure. Physiological and psychological measures were taken during the control, exposure, and recovery periods.

The physiological measures which are considered in this report are mean weighted skin temperature (MWST) and increase in metabolic rate ($\triangle$MR). These are discussed fully in the paper by Iampietro et al. (9) as is the methodology in general. Briefly, MWST represents an average skin temperature based on readings obtained from various parts of the body. Metabolic rate increase, as used herein, refers to the average increase in metabolism (cal./hr.), during exposure, over the reading obtained in the control period, which is an approximate basal reading. The average increase in MR is based on four MR measures taken during the exposure period.

The psychological measure consisted of a paper-and-pencil subjective rating scale on which the *S* indicated how various parts of his body (head, chest, back, arms, hands, legs, and feet) felt on a six-point scale ranging from "very hot" to "very cold." Each *S* indicated how the body parts felt at half-hour intervals during the control, exposure and recovery periods. Only the four exposure ratings are considered in this report.

Results

Agreement between subjects—skin temperature. A MWST reading of each *S* was ob-

tained at the *end* of each two-hour exposure period. The eight conditions were ranked from warmest to coldest for each *S* on the basis of his MWST's. A Kendall coefficient of concordance (12) calculated from the rankings for the six *S*s yielded a *W* of .91 which is significant at better than the .01 level. This indicates that there is a high degree of agreement between the six sets of rankings.

Agreement between subjects—metabolic rate increase. An average △MR for each *S* during each two-hour exposure period was obtained. The eight conditions were ranked from warmest to coldest for each *S* on the basis of the △MRs. A coefficient of concordance calculated from the six sets of rankings yielded a *W* of .83 which is significant at better than the .01 level indicating a high amount of agreement between the six sets of rankings based on MR.

Agreement between subjects—subjective rating scale. Numbers were assigned to the six points on the subjective rating scale, and each *S*'s ratings for all body parts for the four exposure administrations of the scale were averaged. An average subjective response for each *S* for each experimental condition was thus obtained.[2] On the basis of these averages, it was possible to rank the eight conditions for each *S* in terms of how warm or cold they made him feel. The coefficient of concordance calculated from the six sets of rankings yielded a *W* of .76 which is significant at better than the .01 level and indicates a high level of agreement between the six sets of rankings based on subjective ratings.

Comparison of subjective and physiological measures. Having shown that there is a high degree of agreement between *S*s in ranking the eight conditions according to their subjective and physiological responses, it was then permissible to determine the extent of agreement between the subjective and physiological methods of ranking the conditions.

To do this, the individual responses were averaged so that a mean group MWST, △MR, and subjective rating was obtained for each of the eight conditions. The eight conditions were then ranked from "warmest" to "coldest" according to the average measurements. Thus, three sets of rankings of the conditions were obtained; one based on MWST, one on △MR, and one on subjective ratings. It was then possible, using the Spearman rank correlation technique (12), to calculate the degree of agreement between the subjective and physiological methods of ranking the conditions. The group averages and rankings for the MWST, △MR, and subjective rating methods are shown in Table 1. The rank order correlation between the subjective ratings and MWST yielded an r_s of .95 and that between subjective ratings and △MR an r_s of 1.00. These correlations indicate a high degree of agreement between the subjective rating scale and each of the physiological measures in ranking the conditions from warmest to coldest.

Table 1

Comparison of Rankings of Conditions by MWST, ΔMR, and Subjective Methods[a]

Conditions[b,c]	MWST Method	ΔMR Method	Subjective Method
50/30/0	3 (79.8)	5 (85)	5 (1.83)
60/95/0	1 (83.8)	1 (18)	1 (2.73)
50/30/10	7 (69.8)	8 (171)	8 (1.00)
60/30/0	2 (81.3)	2 (38)	2 (2.62)
60/95/10	6 (74.0)	6 (89)	6 (1.73)
50/95/0	4 (79.4)	3 (39)	3 (2.10)
50/95/10	8 (69.4)	7 (119)	7 (1.62)
60/30/10	5 (74.8)	4 (74)	4 (1.85)

[a] The numbers in parentheses refer to the average upon which the ranking is based. MWST averages are in degrees F.; ΔMR averages are in cal./hr.; and subjective averages are based on the 1 through 6 ratings on the subjective scale.
[b] In a condition designation, e.g., 50/30/0, the first figures refer to ambient temperature, the second to percentage of humidity and the third to wind speed.
[c] Conditions are listed in the order in which they were presented.

Study Number Two

Method

Subjects. The *S*s were six enlisted men stationed at the Quartermaster Research & Engineering Command as volunteer test subjects. One *S* had participated in Study Number One.

Procedure. The procedure was exactly the same as described in Study Number One except that the men were clothed in the standard Army fatigue uniform and the ambient temperatures were 40 and 50 degrees F.

[2] The numbers were assigned to the words strictly as a convenience in ordering the data. No assumption was made regarding the equality of the intervals of the subjective scale.

Table 2

Comparison of Rankings of Conditions by MWST, ΔMR, and Subjective Methods

Conditions	MWST Method	ΔMR Method	Subjective Method
50/30/10	5 (75.6)	6 (81)	6 (2.25)
40/95/10	7 (72.0)	8 (101)	8 (1.80)
40/95/0	3 (79.5)	3 (51)	3 (2.83)
50/95/10	6 (74.9)	5 (79)	5 (2.50)
40/30/0	4 (78.8)	4 (59)	4 (2.65)
50/95/0	1 (81.7)	1 (27)	1 (3.08)
50/30/0	2 (80.8)	2 (34)	2 (2.87)
40/30/10	8 (71.4)	7 (93)	7 (2.05)

Results

Agreement between subjects. As in Study Number One, coefficients of concordance were calculated from the rankings based on MWST, △MR and subjective ratings. These were .86, .75, and .85, respectively. All three are significant at better than the .01 level and indicate a high amount of agreement between the six sets of rankings with respect to MWST, △MR, and subjective ratings.

Comparison of subjective and physiological measures. The data were treated the same way as in Study Number One. The comparisons between the MWST, △MR, and subjective rating methods are shown in Table 2. The rank order correlation between subjective ratings and MWST yielded an r_s of .93 and that between subjective ratings and △MR an r_s of 1.00. As in Study Number One, the correlations indicate a high degree of agreement between the subjective rating scale and each of the physiological measures in ranking the conditions from warmest to coldest.

Discussion

The results of both studies indicate that if it is necessary to *compare* environmental conditions and/or order them according to the warmth or coldness of *groups* exposed to them, the subjective rating scale method is as efficient as either of the two physiological methods. The subjective method is also simpler to administer and analyze for scoring.

It should be noted that when actual physiological values are required, the subjective scales are relatively inaccurate in their approximations and cannot replace the physiological measurements.

On the other hand, when an assessment of *Ss'* thermal "feelings" is desired, for other than comparative purposes, the appropriate measure is the subjective one. One cannot accurately infer actual subjective comfort from physiological measurements.

While the two studies did not investigate the comparison of physiological and psychological responses of groups wearing different clothing in a constant environment, the results indicate that the subjective method should be as efficient as the physiological methods in such instances for *comparative* purposes so long as MWST's are not expected to drop below approximately 65 degrees F. There is some evidence that below certain MWST's, the subjective scale loses some of its accuracy.[3] The precise point at which the subjective scale used herein becomes inaccurate has not yet been established.

Summary

Two studies are presented, both of which compare the effectiveness of a subjective rating scale with two physiological measures, mean weighted skin temperature and average increase in metabolic rate, in ranking eight environmental conditions (varying in ambient temperature, humidity and wind speed) from warmest to coldest.

The results of both studies indicate a high degree of consistency between individual responses on each measure and a high degree of agreement between the subjective rating method and both mean weighted skin temperature and metabolic rate methods of ranking the conditions.

It is noted that the subjective scale might be used instead of the physiological measurements for the purpose of *comparing* conditions with regard to the relative warmth or coldness of groups of individuals within the conditions, but *only* for that purpose.

Received January 29, 1958.

[3] John McGinnis, Personal communication. 1957.

References

1. Blair, J. R., Urbush, F. W., & Reed, I. T. Preliminary observations on physiological, nutritional, and psychological problems in extreme cold. Fort Churchill, Canada. Proj. No. 57-3, Field Res. Lab., Ft. Knox, Ky., 1947.
2. Chagas, C., Mortara, G., & Borges Sampaio, F. Alguns estados sobre e sensibilidade termica. *An. Acad. Brazil Cienc.*, 1947, **19**(1), 71–102.
3. Debons, A. Survey of human adjustment problems in the northern latitudes. Personality predispositions of infantrymen as related to their motivation to endure tour in Alaska. A comparative evaluation. Arctic Aeromed. Lab., Ladd AFB, Alaska, 1950. (Proj. Rep. No. 21-01-022.)
4. Ebaugh, F. G., Jr., & Thauer, R. Influence of various environmental stimuli on cold and warmth thresholds. *J. appl. Physiol.*, 1950, **3**, 173.
5. Freedman, A., & Horvath, S. M. Cold weather operations. Test of the adequacy and range of use of winter clothing. Study of methods for selection of men for cold weather operations. Proj. No. 1-1: 1–18, Armored Med. Res. Lab., Fort Knox, Ky., 1944.
6. Hardy, J. D., Goodell, H., & Wolff, H. G. Influence of skin temperature upon pain threshold as evoked by thermal radiation. *Science*, 1951, **114**, 149–150.
7. Hendler, E., & Hardy, J. D. Temperature sensations accompanying changes in skin temperature. NAMC-ACEL-350, Proj. No. NM 1701132, Naval Air Mat. Ctr., Phila., 1957.
8. Herget, C. M., Granath, L. P., & Hardy, J. D. Thermal sensation and discrimination in relation to intensity of stimulus. Am. J. Physiol., 1941, **134**, 645.
9. Iampietro, P. F., Bass, D. E., & Buskirk, E. R. Heat exchanges of nude men in the cold: Effect of humidity, temperature and wind speed. *J. appl. Physiol.*, in press.
10. McCollum, E. L. Personality alteration during reduced caloric intake under survival conditions in the sub-Arctic. Arctic Aeromed. Lab., Ladd AFB, Alaska, 1950. (Proj. Rep., No. 21-01-025.)
11. Patrick, J. R., & Rowles, E. Intercorrelations between metabolic rate, vital capacity, blood pressure, intelligence, scholarship, personality and other measures on university women. *J. appl. Psychol.*, 1933, **17**, 507–521.
12. Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

What happens (*or should happen*) when a supervisor interviews an employee regarding his work . . . ? — READ . . .

THE APPRAISAL INTERVIEW

By NORMAN R. F. MAIER, Professor of Psychology, University of Michigan

This new book offers a penetrating analysis of the cause and effect relations that exist in an interview situation designed to appraise employee efficiency.

The author thoroughly explores the styles of appraisal interviews, each of which requires different skills and is designed to meet different objectives. The first two—the "Tell and Sell" and the "Tell and Listen" methods—deal with employee development. The third method—"Problem-Solving"—is concerned with job improvement.

The Appraisal Interview describes the problems faced by the interviewer, the skills needed to make the interview successful by overcoming these problems, and the importance of clarifying the objectives of the interview in order to reach a productive conclusion. The dangers involved are pointed out, showing how an interview, if improperly handled, can do more harm than good. The book shows why interviews go wrong and what can be done to save them. Actual transcriptions of interviews, conducted by different methods, have been included to illustrate specific problems. 1958. 246 pages. $5.95.

Send for an examination copy.

JOHN WILEY & SONS, Inc., 440 Fourth Avenue, New York 16, N.Y.

PSYCHOLOGICAL MONOGRAPHS

A sample of monographs still in print:

Factors Related to Success in the Human Relations Aspect of Work-Group Leadership, by H. H. MEYER. No. 320. *$1.00*

Occupational Aptitude Patterns of Selected Groups of Counseled Veterans, by W. L. BARNETTE, JR. No. 322. *$1.50*

The Academic Achievement of Veteran and Non-Veteran Students, by NORMAN FREDERIKSEN and W. B. SCHRADER. No. 347. *$1.50*

Errors in Time-Study Judgments of Industrial Work Pace, by K. A. LIFSON. No. 355. *$1.00*

The Prediction of Effectiveness as a Factory Worker, by M. M. BRUCE. No. 362. *$1.00*

Job Satisfaction as Related to Need Satisfaction in Work, by R. H. SCHAFFER. No. 364. *$1.00*

Improving the Employability and Attitudes of "Difficult to Place" Persons, by ALFRED FEINTUCH. No. 392. *$1.00*

Order now before these monographs go out of print.

Write for complete list of items still available.

American Psychological Association
1333 Sixteenth Street, N. W.
Washington 6, D. C.